

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:20:24 ; Search time 31.79 Seconds
(without alignments) (without alignments)
1046.951 Million cell updates/sec

Title: US-09-189-415A-2

Perfect score:

Sequence: 1 MPIGNLGNNVNGNHLIPPAP.....GETAVSSVNAAPTGPVRFV 549

Scoring table: BLOSUM62

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

| Minimum | DB seq | length: |
|---------|--------|--------------------|
| Maximum | DB seq | length: 2000000000 |

Post-processing: Minimum Match 0%

Fast processing: Minimum Match 0%
Maximum Match 100%

Database :

| | |
|-----|--|
| 1: | /SDSI1/gcgdata/geneseq/geneseqp/AA1980.DAT:* |
| 2: | /SDSI1/gcgdata/geneseq/geneseqp/AA1981.DAT:* |
| 3: | /SDSI1/gcgdata/geneseq/geneseqp/AA1982.DAT:* |
| 4: | /SDSI1/gcgdata/geneseq/geneseqp/AA1983.DAT:* |
| 5: | /SDSI1/gcgdata/geneseq/geneseqp/AA1984.DAT:* |
| 6: | /SDSI1/gcgdata/geneseq/geneseqp/AA1985.DAT:* |
| 7: | /SDSI1/gcgdata/geneseq/geneseqp/AA1986.DAT:* |
| 8: | /SDSI1/gcgdata/geneseq/geneseqp/AA1987.DAT:* |
| 9: | /SDSI1/gcgdata/geneseq/geneseqp/AA1988.DAT:* |
| 10: | /SDSI1/gcgdata/geneseq/geneseqp/AA1989.DAT:* |
| 11: | /SDSI1/gcgdata/geneseq/geneseqp/AA1990.DAT:* |
| 12: | /SDSI1/gcgdata/geneseq/geneseqp/AA1991.DAT:* |
| 13: | /SDSI1/gcgdata/geneseq/geneseqp/AA1992.DAT:* |
| 14: | /SDSI1/gcgdata/geneseq/geneseqp/AA1993.DAT:* |
| 15: | /SDSI1/gcgdata/geneseq/geneseqp/AA1994.DAT:* |
| 16: | /SDSI1/gcgdata/geneseq/geneseqp/AA1995.DAT:* |
| 17: | /SDSI1/gcgdata/geneseq/geneseqp/AA1996.DAT:* |
| 18: | /SDSI1/gcgdata/geneseq/geneseqp/AA1997.DAT:* |
| 19: | /SDSI1/gcgdata/geneseq/geneseqp/AA1998.DAT:* |
| 20: | /SDSI1/gcgdata/geneseq/geneseqp/AA1999.DAT:* |
| 21: | /SDSI1/gcgdata/geneseq/geneseqp/AA2000.DAT:* |
| 22: | /SDSI1/gcgdata/geneseq/geneseqp/AA2001.DAT:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query % | | Length | DB | ID | Description |
|------------|-------|---------|------|--------|----------|--------------------|-------------|
| | | Match | | | | | |
| 1 | 2798 | 99.9 | 549 | 20 | AAV06220 | EPEC E. coli trans | |
| 2 | 1492 | 53.3 | 559 | 20 | AAV06221 | EHEC E. coli trans | |
| 3 | 530 | 18.9 | 107 | 21 | AAB20576 | Intimin C-terminal | |
| 4 | 165 | 5.9 | 30 | 20 | AAV06213 | EPEC E. coli trans | |
| 5 | 149 | 5.3 | 831 | 19 | AAW68201 | M. catarrhalis str | |
| 6 | 141 | 5.0 | 580 | 18 | AAW10660 | Endoglycoceramidas | |
| 7 | 139.5 | 5.0 | 2411 | 21 | AAB23863 | Haemophilus influe | |
| 8 | 138.5 | 4.9 | 2353 | 17 | AAV99393 | Haemophilus adhesi | |
| 9 | 136.5 | 4.9 | 3096 | 22 | AAB46770 | FANCI3 protein fr | |
| 10 | 136.5 | 4.9 | 4630 | 18 | AAW19629 | Streptomyces venez | |
| 11 | 135 | 4.8 | 1532 | 21 | AAB40945 | Human ORF709 | |

| | | | | | | |
|----|-------|-----|------|----|-----------|---------------------|
| 12 | 133.5 | 4.8 | 1978 | 20 | AAV27230 | Amino acid sequenc |
| 13 | 132.5 | 4.7 | 1026 | 18 | AAV37490 | Caulobacter cresce |
| 14 | 132.5 | 4.7 | 1026 | 21 | AAV44757 | Caulobacter cresce |
| 15 | 132.5 | 4.7 | 4630 | 21 | AAV71717 | S. venezuelae vep |
| 16 | 131.5 | 4.7 | 1026 | 15 | AAAR4893 | rsaA S-lyaaer prote |
| 17 | 130.5 | 4.7 | 1026 | 17 | AAV94014 | Caulobacter S-layer |
| 18 | 129 | 4.6 | 1341 | 21 | AAV85657 | Human Acinus L pro |
| 19 | 128.5 | 4.6 | 1536 | 18 | AAW30293 | Non-typeable Haemo |
| 20 | 128 | 4.6 | 850 | 19 | AAWG2040 | Protein isolated f |
| 21 | 128 | 4.6 | 1237 | 18 | AAW55640 | H. pylori ORF 04ep |
| 22 | 128 | 4.6 | 1237 | 20 | AAV17187 | H. pylori outer me |
| 23 | 126 | 4.5 | 376 | 19 | AAW81744 | M. tuberculosis im |
| 24 | 126 | 4.5 | 376 | 19 | AAW64377 | Mycobacterium tube |
| 25 | 126 | 4.5 | 376 | 20 | AAV39174 | M. tuberculosis an |
| 26 | 126 | 4.5 | 376 | 20 | AAV39031 | M. tuberculosis re |
| 27 | 125.5 | 4.5 | 1354 | 21 | AAV74857 | Neisseria gonorrhe |
| 28 | 124.5 | 4.4 | 2514 | 21 | AAV75097 | Neisseria meningit |
| 29 | 124 | 4.4 | 1180 | 21 | AAAB01845 | Haemophilus influe |
| 30 | 124 | 4.4 | 1188 | 21 | AAAB01844 | Haemophilus influe |
| 31 | 124 | 4.4 | 1536 | 15 | AAAR63505 | Haemophilus high m |
| 32 | 124 | 4.4 | 1536 | 21 | AAAB01846 | Haemophilus influe |
| 33 | 124 | 4.4 | 2541 | 21 | AAAB1087 | Human ORFX ORF851 |
| 34 | 123.5 | 4.4 | 3596 | 21 | AAV67407 | Bordetella pertuss |
| 35 | 123.5 | 4.4 | 3647 | 11 | AAAR05041 | Filamentous haemag |
| 36 | 123 | 4.4 | 741 | 9 | AAAP80136 | Neisseria IgA-prot |
| 37 | 121 | 4.3 | 608 | 19 | AAAV79280 | Arabidopsis phytoc |
| 38 | 121 | 4.3 | 1638 | 20 | AAAY00138 | Enterococcus faeca |
| 39 | 121 | 4.3 | 1638 | 20 | AAAY00140 | Enterococcus faeca |
| 40 | 121 | 4.3 | 1638 | 20 | AAAY00142 | Enterococcus faeca |
| 41 | 121 | 4.3 | 2042 | 19 | AAW56319 | Haemophilus paraga |
| 42 | 120.5 | 4.3 | 979 | 21 | AAAB10763 | Murine T protein f |
| 43 | 120.5 | 4.3 | 2039 | 19 | AAW56332 | Haemophilus paraga |
| 44 | 120 | 4.3 | 1048 | 18 | AAW27277 | Human cytomagalovi |
| 45 | 119.5 | 4.3 | 1721 | 19 | AAW52847 | A. mediterranei ri |

ALIGNMENTS

| | | |
|-----------|------------|---|
| RESULT | 1 | |
| AAAY06220 | | |
| ID | AAAY06220 | standard; Protein; 549 AA. |
| XX | AC | |
| XX | AAAY06220; | |
| XX | DT | 16-AUG-1999 (first entry) |
| XX | XX | EPEC E. coli translocated intimin receptor (Tir). |
| DE | DE | Tir; translocated intimin receptor; Hp90; enteropathogenic; |
| XX | XX | EPEC; infection; diagnosis; vaccine. |
| KW | XX | Escherichia coli. |
| OS | OS | |
| XX | Key | Location/Qualifiers |
| XX | FT | Domain |
| XX | FT | 234..253 |
| XX | FT | /note= "putative transmembrane domain" |
| XX | FT | 364..386 |
| XX | FT | /note= "putative transmembrane domain" |
| XX | FT | Misc-difference 180 |
| XX | FT | /note= "encoded by AAA" |
| XX | FT | Misc-difference 314 |
| XX | FT | /note= "given as xaa in the specification; Lys |
| XX | FT | is deduced from the DNA sequence" |
| XX | FT | |

| | |
|----|---------------|
| XX | WO924576-A1. |
| PN | 20-MAY-1999. |
| XX | |
| XX | |
| PD | |
| XX | |
| PF | 10-NOV-1998; |
| XX | 98WO-CA01042 |
| XX | |
| PR | 12-NOV-1997; |
| XX | 97US-0065130. |

QY 481 GYSVIONFSGSPVTRGLTGTGQGIQSTYALLANSGLRLMGGLTSGGETAVSSVNA 540
Db 481 gysvionfsgspvtrglgtgqgiqstyalansgllrglmggltsgetavssvnaa 540
QY 541 PTPGPVRFV 549
Db 541 ptpgpvrfrv 549

RESULT 2
AAY06221
ID AAY06221 standard; Protein; 559 AA.
AC AAY06221;
XX
DT 16-AUG-1999 (first entry)
XX
DE EHEC E. coli translocated intimin receptor (Tir).
XX
KW Tir; translocated intimin receptor; Hp90; enterohaemorrhagic;
KW EHEC; infection; diagnosis; vaccine.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT Misc-difference 453 /note= "encoded by codon of 1 apparent nucleotide,
FT causing frameshift in the DNA sequence"
XX
PN W09924576-A1.
XX
PD 20-MAY-1999.
XX
PF 10-NOV-1998; 98WO-CA01042.
XX
PR 12-NOV-1997; 97US-0065130.
XX
XX (UYBR-) UNIV BRITISH COLUMBIA.
XX
XX Devinney R, Finlay BB, Kenny B, Stein M;
XX
XX WPI: 1999-337712/28.
XX
XX N-PSDB; AAX58859.

New translocated intimin receptor useful for treating infection by
enteropathogenic or enterohaemorrhagic Escherichia coli
Claim 7; Page 55-58; 91pp; English.

The present sequence represents Tir, a novel translocated intimin
receptor (formerly termed Hp90) from an enterohaemorrhagic
Escherichia coli (EHEC) strain. The sequence was deduced from an
isolated tir polynucleotide (see AAX58859). Tir proteins are
secreted by attaching and effacing pathogens such as EHEC and EPEC
(see AAY06220) E. coli. The bacterial pathogens insert their own
receptors into mammalian cell surfaces, to which the pathogen then
adheres to trigger additional host signaling events and actin
nucleation. Diagnosis of disease caused by pathogenic E. coli can
be performed by use of antibodies that bind to Tir to detect the
protein or the use of nucleic acid probes for detection of nucleic
acids encoding Tir polypeptide. Isolated Tir nucleic acids, Tir
peptides, a recombinant method for producing recombinant Tir,
antibodies which bind to Tir, and a kit for the detection of
Tir-producing E. coli are provided. A method of immunising a host
with Tir to induce a protective immune response is also provided.
In addition, Tir fusion proteins can be used in attenuated E. coli
to induce a cell-mediated immune response to other polypeptides,
e.g. antigens. A method for screening for compounds which
interfere with the binding of bacterial pathogens to their
receptors is further provided.

Sequence 559 AA;

PA (UYBR-) UNIV BRITISH COLUMBIA.
XX Devinney R, Finlay BB, Kenny B, Stein M;
XX
XX WPI: 1999-337712/28.
XX
XX N-PSDB; AAX58859.

XX New translocated intimin receptor useful for treating infection by
PT enteropathogenic or enterohaemorrhagic Escherichia coli

XX Claim 6; Page 55-58; 91pp; English.

XX The present sequence represents Tir, a novel translocated intimin
CC receptor (formerly termed Hp90) from an enteropathogenic
CC Escherichia coli (EPEC) strain. The sequence was deduced from an
CC isolated tir polynucleotide (see AAX58859). Tir proteins are
CC secreted by attaching and effacing pathogens such as EPEC and EHEC
CC (see AAY06221) E. coli. The bacterial pathogens insert their own
CC receptors into mammalian cell surfaces, to which the pathogen then
CC adheres to trigger additional host signaling events and actin
CC nucleation. Diagnosis of disease caused by pathogenic E. coli can
CC be performed by use of antibodies that bind to Tir to detect the
CC protein or the use of nucleic acid probes for detection of nucleic
CC acids encoding Tir polypeptide. Isolated Tir nucleic acids, Tir
CC peptides, a recombinant method for producing recombinant Tir,
CC antibodies which bind to Tir, and a kit for the detection of
CC Tir-producing E. coli are provided. A method of immunising a host
CC with Tir to induce a protective immune response is also provided.
CC In addition, Tir fusion proteins can be used in attenuated E. coli
CC to induce a cell-mediated immune response to other polypeptides,
CC e.g. antigens. A method for screening for compounds which
CC interfere with the binding of bacterial pathogens to their
CC receptors is further provided.

XX: Sequence 549 AA;

Query Match 99.9%; Score 2798; DB 20; Length 549;
Best Local Similarity 99.8%; Pred. No. 1.1e-210;
Matches 548; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPIGNLNNVNGNHLIPAPPSPDTCGAARGTGHLISSTGALGSRSLFSPRLNSMADS 60
Db 1 mpignlgnnvngnhlppappspdgdaargtgthllsstgalgsrlfsgprlnsmads 60
QY 61 VDSRDIPGLPTNPRLAATSTCLLGGFVLHDKGDLILNTQIGPSAFRVEVQADGTH 120
Db 61 vdsrdipglptnpsrlaaatsetcll9gfevlhdkgpldiltntqigpsafrevevqadgth 120
QY 121 AATGKNGLEVSTLSPQWSSLSQSTIDTEGKNRFVFTGGGSGHPMVTVASDIAEARTR 180
Db 121 aatgknglevstlspqewsslsqstidtegnrfvftggrgsgghpmvtvasdiaeart 180
QY 181 ILAKLPDNGHGRQKDVTRSVGVSGASIDGIVVSEHTTSTNVSRSRDPKFWVSGA 240
Db 181 ilaklpdnhgrqkdvtrsvgvsgasidgvgvsehtstntnvsrsrdbpkfwvsga 240
QY 241 IAAGLAGLAATGTAQALATPEPDDPTTDDPQANAASATKDLQTOAFKNPENQKVN 300
Db 241 iaaglaglaatgtaqalaltpepddpttdpqaanaasatkdlqtqafknpenqkvn 300
QY 301 IDANGNAIPSGELXDDIVEIOIAQAKEAGEVARQQAQVESAQAQRYEQAARRQELOL 360
Db 301 idangnaipsgelkddiveioiaqakeagevarqqaqvesaqaqrqedqharrqeelql 360
QY 361 SSGIGVGLSALLVAGGIGAGVTALHRRNPQAEQTTTTHFVQQQGGIPQHKVALLM 420
Db 361 ssgigvllsallvaggigagvtalhrnpqaeqtthfvlvqqgqg99ipqhkvalm 420
QY 421 PQRRRFRDRDQSGSVASTHWSDDSEVVNPNYAEGVGGARNLSAHPQEEHIYDEVAADP 480
Db 421 pqrrrfrdrdsgsvasthwsddsevvnpnyaevggarnlsahpqeehiydevaadp 480

Query Match 53.3%; Score 1492; DB 20; Length 559;
Best Local Similarity 54.7%; Pred. No. 1.7e-108;
Matches 322; Conservative 64; Mismatches 133; Indels 70; Gaps 12;

QY 1 MPIGNLNNVNGNHLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
DB 1 mpignlghnpvnnslppappplpsqtdga--ggrgqlinstgplgsralftprvnsmds 58

QY 61 VDSR--DIPGLTFNPSRLAAATSECLLGGFEVLHDKGLDILNTQIGPSAFRVEQADG 118
DB 59 gdnrasdvpglbvpnmrlaa--seitlndgfevlhdkgpldlnrqigsvfrvetqdg 116

QY 119 TAAAGEKNGLEVSVTLSPOEWSLQSDTEKGNRFVFTGGRGGSGHPMVTVASDIAEAR 178
DB 117 khlaavgqngvsvslsdeqeyarqsldepegkdfvftggrggghamvtvasditear 176

QY 179 TRILAKLPDNNHGRQPKDVTDRSVGVGSASGI-----DDGV--VSEPHTTTSSVRS 230
DB 177 qrillelepkgtg-----eskgageskgvgelresnsgaenttetqtststslrs 227

QY 231 DKFWVSVCALAGLAGLAATGIAQALATPEPDPPTTDPQANAASATKDLTQEA 290
DB 228 dplkwlalgtvatgliaatgivaalaitpepdspttdpdaasatetatrdqltkea 287

QY 291 FKPNENOKVNDANGNAIPSGELXDDIVBQIAQAKEAGEVARQQAQVESNAQAQRYEDQ 350
DB 288 fgnpdnqknidelnaiapsvylkddvvanieeqkaageakqaieannaqakkydeq 347

QY 351 HARQEEQLSSGIGYGLSSALIVAGGAGVYTTALHRRNQPAEQTTTTT-----HT 403
DB 348 qakrqeelkvssgagylsgalllggigvavtaalhrknqveqtttttttttsart 407

QY 404 VVQQQTGGTQPHKVALMPQERRFRSDRRSQCSVASVTHWSDSSEVVNPYAEVGG---AR 460
DB 408 venkpanntpcagqnvtdpgsedtmesrrssmaststffdtss-----lpgpccrir 458

QY 461 NLSLSAHO-----PEEHYDEVAADPGSVIQNFGSGGPVTGRLIG 500
DB 459 mmlmkhrcmirrcrlllrlfrwigigisvvytiqhpp-----rdtdtng---arlig 510

QY 501 TPCQGGTQSYALLANSGLGURLGMGGTSGGETAVSSVNAAPTGPGRFV 549
DB 511 npsagiqtayarlalsgglrhdmgiltggsnsavntsnppagpghrfv 559

RESULT 3
AAB20576
XX AAB20576 standard; protein; 107 AA.
AC AAB20576;
XX AAB20576;
DT 08-DEC-2000 (first entry)
XX DE Intimin C-terminal Tir binding domain amino acid sequence.
XX DE Intimin; Tir binding domain; inhibitor; intimin adhesion; screening;
KW Tir-independent eukaryotic cell binding activity; bacterial infection;
KW diarrhoea; antibacterial.
OS Unidentified.
XX OS
XX WO200045173-A1.
PN WO200045173-A1.
XX PD 03-AUG-2000.
XX PF 31-JAN-2000; 2000WO-GB00254.
XX PR 29-JAN-1999; 99GB-0001897.
XX PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX PA Frankel GM, Matthews SJ, Hale CB, Dougan G;
XX PI

XX WPI; 2000-499357/44.
XX Screening for inhibitors of intimin binding to eukaryotic cells, for
PT use in diagnosing, preventing and treating bacterial infections,
PT especially *Escherichia coli* O157:H7 -
XX Claim 8; Page 76; 96pp; English.
XX The present invention describes a method of screening for an inhibitor
CC of intimin binding to eukaryotic cells. The method comprises exposing an
CC intimin polypeptide having a Tir-independent cell binding activity to
CC test agents, and obtaining an inhibitor based on its ability to bind the
CC polypeptide. The inhibitors are used in the prevention, treatment and/or
CC diagnosis of bacterial infections, preferably by enteropathic and/or
CC enterohaemorrhagic *Escherichia coli*, *Shiga* toxinigenic *E. coli*, *Hafnia*
CC *alvei* or *Citrobacter freundii*, or especially *E. coli* O157:H7. The
CC infections cause a histopathological effect known as attachment and
CC effacement on intestinal epithelial cells. The inhibitors can be used
CC to produce food supplements or additives, especially where the food is
CC a milk substitute. The method can be used to sort cells based on their
CC ability to bind to a Tir independent cell binding domain of an intimin
CC polypeptide. Polypeptides having Tir-independent intimin binding
CC activity can be used to produce a vaccine against a bacterial disease.
CC The present sequence represents a specifically claimed intimin C-terminal
CC Tir binding domain amino acid sequence, for use in the method of the
CC present invention.
XX Sequence 107 AA;
SQ

Query Match 18.9%; Score 530; DB 21; Length 107;
Best Local Similarity 99.1%; Pred. No. 3.1e-34;
Matches 106; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 255 QALALTPEDPTTDPQANAASATKDLTQEAFAKNPENQKVNIDANGNAIPSGELX 314
DB 1 qalaltpepdpdttdpqaanaaesatkdlqtqeaafknpenqkvnidangnaipsgeik 60

QY 315 DDIVQIAQAQAKEAGEVARQQAQVESNAQAQRYEDQHAHQEELQLS 361
DB 61 ddiveqiaqakeagevarqqaqavesnaqagqrqedgharreeelqls 107

RESULT 4
AAY06213
ID AAY06213 standard; Peptide; 30 AA.
XX AAY06213;
AC AAY06213;
DT 16-AUG-1999 (first entry)
XX DE EPEC *E. coli* translocated intimin receptor N-terminal peptide.
XX DE Tir; translocated intimin receptor; Hp90; enteropathogenic;
KW EPEC; infection; diagnosis; vaccine.
XX OS *Escherichia coli*.
XX OS
XX WO9924576-A1.
PN WO9924576-A1.
XX PD 20-MAY-1999.
XX PF 10-NOV-1998; 98WO-CA01042.
XX PR 12-NOV-1997; 97US-0065130.
XX PA (UYBR-) UNIV BRITISH COLUMBIA.
XX PA Devinney R, Finlay BB, Kenny B, Stein M;
XX WPI; 1999-337712/28.
XX

PT New translocated intimin receptor useful for treating infection by
PT enteropathogenic or enterohemorrhagic Escherichia coli
XX
XX
XX Example 1; Page 37; 91pp; English.
XX
XX The present sequence represents the N-terminal sequence of Tir (see
CC also AAY0620), a novel translocated intimin receptor from an
CC enteropathogenic Escherichia coli (EPEC) strain. The 78 kDa EPEC
CC protein is secreted by the bacterial pathogen. Diagnosis of
CC disease caused by pathogenic E. coli can be performed by use of
CC antibodies that bind to Tir to detect the protein, or the use of
CC nucleic acid probes for detection of nucleic acids encoding Tir. A
CC kit for the detection of Tir-producing E. coli is provided. Also
CC provided are a method of immunising a host with Tir to induce a
CC protective immune response, and a method for screening for
CC compounds which interfere with the binding of bacterial pathogens
CC to their receptors.
XX
XX
SQ Sequence 30 AA;

Query Match 5.9%; Score 165; DB 20; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 PIGNLGNVNGNHLIPAPPPLPSQTGGAAR 31
DB 1 pignlgnvngnhlipappplpsqtdgaar 30

RESULT 5
AAW68201
ID AAW68201 standard; Protein; 831 AA.
XX:
AC AAW68201;
XX
XX 07-OCT-1998 (first entry)
XX M. catarrhalis strain O35E UspA1 antigen.
XX Moraxella catarrhalis; UspA1; UspA2; antigen; genetic vaccination;
KW vaccine; otitis media; sinusitis; lower respiratory tract infection;
KW immunity enhancer; immunoassay reagent.
XX
XX Moraxella catarrhalis.
XX
XX WO9828333-A2.
XX
XX 02-JUL-1998.
XX
XX 19-DEC-1997; 97WO-US23930.
XX
XX 20-DEC-1996; 96US-0033598.
XX (TEXA) UNIV TEXAS SYSTEM.
XX Aebi C, Cope LD, Fiske MJ, Fredenburg R, Hansen EJ;
PI Maciver I;
XX
XX WPI; 1998-377595/32.
DR N-PSDB; AAV41341.
XX
XX New peptide(s) containing the core epitope of Moraxella catarrhalis
PT Usp proteins - useful in, e.g. vaccines to prevent or treat M.
PT catarrhalis infection, and antibodies for passive immunisation.
XX
XX Claim 22; Pages 133-136; 237pp; English.
XX This represents a UspA1 antigen of Moraxella catarrhalis strain O35E.
CC Nucleic acid sequences encoding the UspA1 and A2 antigens of
CC M. catarrhalis isolates O35E, O46E, TTA24 and TTA37 can be used in
CC genetic vaccination. An antigenic composition or vaccine containing
CC antigenic peptides from UspA1 or UspA2 antigens are used to induce an

CC immune response in mammals against M. catarrhalis and can be used to
CC treat infections such as otitis media, sinusitis, lower respiratory
CC tract infections. They can also be used as immunity enhancers for other
CC bacterial, parasitic or viral antigens, to raise antibodies and as
CC immunoassay reagents for detecting specific antibodies. The antibodies
CC are useful for passive immunisation and as immunoassay reagents.
CC Detection of the epitopic core sequence, by immunoassay or by PCR, is
CC used to diagnose infection. The usp antigens encoding nucleic acid
CC sequences are also used to produce recombinant proteins and for screening
CC for potential anti-M. catarrhalis agents, while their fragments are
CC useful as diagnostic probes or primers or to isolate variant sequences.
XX
XX Sequence 831 AA;

Query Match 5.3%; Score 149; DB 19; Length 831;
Best Local Similarity 19.2%; Pred. No. 0.0033;
Matches 101; Conservative 73; Mismatches 193; Indels 160; Gaps 21;
QY 74 SRLAAATSETCLLGGFEVLHDKGPLDLNTQIGP---SAFRVEVQADGTHAIGEKNGLE 130
DB 226 sgttgvtstnsvllgneta--gkqattvknaevgglsaltgfageskaevgsvgseggr 283
QY 131 VSYTLSPQEWSSLOSIDTEGKNRFVFTGGGGSGHPMVTVAS----DIAEARTRILAKLD 186
DB 284 qivnvga---ggisdstsdavn-----gsqhalatvvdnqydivnnradilnnqd 332
QY 187 P-----DNHGGROPKVDVT-----RSVGVS----- 207
DB 333 dikhlcqevkgldnevgelsrdinslhvtdnqddikelkrgvkeldnevgvlrsdins 392
QY 208 -----ASGIDDGVVSEHTSTTNSVRSOPKFWVSGATAAGLAGLAATGIAQALALTP 262
DB 393 lhadvadnqddiaknkadikglkvevkeldk---evgvisrdigsl----- 435
QY 263 PDPPTTTDPQAAANAASATKDQLTQEAENPE-----NOKVNIDAN----- 304
DB 436 hadvatnqadiaknqadiktlennveeellnlsgrlldqkadidnnlnniyelaqqdqh 495
QY 305 -----GNAIPSG--ELXDDIVEQIAQAKEAGEVARQAVESNAQAQOQRYEDQHARRQ 355
DB 496 ssdiktiknnveeglldslgrlldqkadiknqadiagnqtdldlaaynelqdyqakq 555
QY 356 EE-----LQLSSGIG--YCL-----SSALIVAGGIGAGVTTAL 386
DB 556 teaidalnassentqniaknqadianniniyelaqqdqhssdiktakvsaantdri 615
QY 387 HRRNQPAE--QTTTTHVVOQTGGIPQHKVALMPQRRRFRSRRDRSQGSVASTHWS 443
DB 616 aknkaeasafetltknqntliedqealvegnka--inqelegfaahadiq----- 664
QY 444 DSSSEVNPYAEVCGARNSLSAHQPEEHYDEVAADPGYSVIONFSG 490
DB 665 --dkqilqnqadiittktai-----eqnairtva--ngfelekknag 702

RESULT 6
AAW10660
ID AAW10660 standard; Protein; 580 AA.
XX
XX AAW10660;
XX
XX 20-JUL-1997 (first entry)
XX Endoglycoceramidase activator II.
XX
XX Endoglycoceramidase; activator; glycolipid; enzyme.
XX Rhodococcus spp. M-777.
XX
XX EP759470-A2.
XX
XX 26-FEB-1997.
PD

XX 28-JUN-1996; 96EP-0110513.
XX 29-JUN-1995; 95JP-0188466.
XX (TAKI) TAKARA SHUZO CO LTD.
XX Ito M, Izu H, Izumi Y, Kato I, Kurome Y, Sano M;
XX WPI; 1997-147519/14.
XX N-PSDB; AAT61050.
XX DNA encoding endo:glyco:cer:amidase activator polypeptide - for
XX prodn. of recombinant polypeptide, useful in glyco:lipid analysis
XX Claim 1; Page 13-15; 30pp; English.
XX DNA encoding endoglycoceramidase activator polypeptide can be used
XX for the prodn. of the recombinant polypeptide, which can be used
XX to research the intracellular function of glycolipids, because
XX it shifts the optimum pH for endoglycoceramidase II towards
XX neutral, allowing the latter to be used to hydrolyse glycolipids
XX even at pH 7.5.
XX Sequence 580 AA;

Query Match 5.0%; Score 141; DB 18; Length 580;
Best Local Similarity 23.9%; Pred. No. 0.0085;
Matches 84; Conservative 45; Mismatches 126; Indels 96; Gaps 16;
QY 4 GNLGNNVNGNHLIPP-----APLPSTQDGAARGGTGHLISSTGALGSRSLFSP 53
DB 215 gqfndanfllflprasapivgtwaptqcpdrtaaa-----gplnagagplatiqi---l 267
QY: 54 RNSMADSVDSRDPGLPNPSR--LAAATSEFCLLGFEVLHD---KGPLDILNTQIGP 107
DB 268 rqaava-tvsyldgpsavtnggfeftinatvptpdsgqvqftrdgedgapvdlvn---gk 323
QY 108 SAFRVEVQADGTHAAGEKNGLEV-----SVTLSPQEWSSLSIDTEGKNRFVFGG 159
DB 324 asltsgldtdgdgyaekflgaefnfpssaaktvtsgdittsvt----- 371
QY 160 RGGSGHPMVTVASDIAEARTRLAKLDPNHHGGRQPKVDTRSVGVSGASGIDGWWSET 219
DB 372 ---gpdh-----dayrdpvnltakvepgvsggtvafevd--gtpvgtdavmdmdgaavlp 421
QY 220 HTSTTNSSVRSDPKFWVSVGAIAAGLAGLAATGIAQALAL-----TPEDDPTTTTDP 271
DB 422 htfttngthr-----viarysg--aegispsvslqypvsvteapaadvattitv 468
QY 272 DQANNAESATKDQLTQFAFKNPENQK-----VNIDANGNA 307
DB 469 dplastakgs---pvtlartldpadargvtqfklgdvllgpgvrvdangva 516

RESULT 7
AAB23860
ID AAB23860 standard; Protein; 2411 AA.
XX AAB23860;
XX AAB23860;
XX 17-JAN-2001 (first entry)
XX Haemophilus influenzae adhesin (Hia) protein from type c strain API.
XX Hia; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine;
XX non-typeable Haemophilus influenzae; antiinflammatory; auditory;
XX antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
XX diagnosis; immunogenic; antigen.
XX Haemophilus influenzae.

PN WO200055191-A2.
XX 21-SEP-2000.
XX 16-MAR-2000; 2000WO-CA00289.
XX 16-MAR-1999; 99US-0268347.
XX (CONN-) CONNAUGHT LAB LTD.
XX Loosmore SM, Yang Y, Klein MH;
XX WPI; 2000-618897/59.
XX N-PSDB; AAA92499.
XX Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
XX use as antigens and vaccines and for treating Hemophilus influenzae
XX infection -
XX Claim 1; Fig 24; 275pp; English.
XX The present sequence represents a Haemophilus influenzae adhesin
XX (Hia) protein from the type c Haemophilus influenzae strain API.
XX Hia genes and proteins have antiinflammatory, auditory and antibacterial
XX activities, and can be used in the production of a vaccine. An
XX immunogenic composition comprising an Hia gene, a polypeptide encoded
XX by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
XX protection against disease caused by Haemophilus strains in a
XX susceptible host, preferably a human. An Hia protein is useful as an
XX antigen, in immunogenic preparations including vaccines, as a carrier
XX for other immunogens, and in the generation of diagnostic reagents. Hia
XX is useful for treating diseases caused by the infection of Haemophilus
XX influenzae such as meningitis, epiglottitis, septicaemia and otitis
XX media. Recombinant production of Hia favours high recovery of the
XX protein compared to the low recovery of native protein from Haemophilus
XX influenzae species. A truncated protein has a significantly higher
XX amount of recovery than a full-length protein.
XX Sequence 2411 AA;

Query Match 5.0%; Score 139.5; DB 21; Length 2411;
Best Local Similarity 21.5%; Pred. No. 0.079;
Matches 125; Conservative 83; Mismatches 228; Indels 145; Gaps 28;
QY 6 LGNNVNGNH---LIPPAPPLPSQTDGAARG---GTGHLISST--GALGSRSLFSPLRNS 56
DB 1161 lggtagrndrtgtvinkdgltitlangaaagtdasngntisvtkdgisagnkeintvksa 1220
QY 57 MADSVDSRDPGLPNPSRLAAATSETCLLGFEVLHDKGPLDILNTQIGPSAFRVEVOA 116
DB 1221 lkytdtqntag-atqp---aantaeva-----kqdlvdltkpatgaagnadaka 1267
QY 117 -DGHAAAGEKNGL-----EVSVTLSPOEW-----SSLSQSIDTEGKNRFVFTGGRGGSG 164
DB 1268 pdttaatvgdlrglglwlsakktadettqkfehaavknevefvgkngatvsaktddng 1327
QY 165 HPMVTVASDIAEAR-----TRILAKLDPNHHGGRQPKVDTRSVGVSGASGIDGWW 216
DB 1328 khtvti--dvaekvgdglektdgkikikvndtgnlltvdadtgasvakgefnavt 1384
QY 217 SEHTST-TNSSVR-----SDPKFWVSVGAIAAGLAGLAATGIAQALALT 260
DB 1385 tdatagtnanergkvkvkngngatattetdkkvtatvgdvakaind-aatfv-----k 1437
QY 261 PEPDPTTTTDPQANNAESATK--DQLTQFAFKNPENQKVNID-----ANGNAIPS 310
DB 1438 venddsatiddspdtgdandalkagdtltclikagk---nlkvrdgknitfalandsvks 1494
QY 311 GELXDDIVEQIAQQAKEAGEVARQQAQRYEDQHARRQEEQLSSGIGYGLSS 370
DB 1495 atvsdkl-----slgtngknknitsdtkglnfakdsktgddanhl-ngiastltd 1544

QY 371 ALIVAGGIGAVTTALHRRNQPAEQTTTTTHTVVOOQTGGIPQHKVALMPQERRRFSR 430
Db 1545 tl-----lsgattnl-----gg-----ngitdnekkraasv 1571
QY 431 RDSQGSVASTHWSDSSESSEVNPYAEVGGARNSLSAHOPEHIYDEVAADPGYSVIONFSG 490
Db 1572 kd-----vlnagvnrvvkpa-sannqveni-dfvt---ydtvdfvsg 1610
QY 491 SGPVTRGLRGTPOGQISTYVALLANGGLRLMGGLTSGGE 531
Db 1611 dkdtstvtveskdngkrtevkigaktsvikdhngkiftgke 1651

RESULT 8
AAR99393
ID AAR99393 standard; Protein; 2353 AA.
AC AAR99393;
XX 15-JAN-1997 (first entry)
DT Haemophilus adhesion protein HA2.
DE Haemophilus adhesion protein; HA2; hsf protein; vaccine.
KW Haemophilus influenzae type b strain C54.
OS Haemophilus influenzae type b strain C54.
XX WO9630519-A1.
XX 03-OCT-1996.
XX 22-MAR-1996; 96WO-US04031.
XX 24-MAR-1995; 95US-0409995.
PR (UYSL-) UNIV ST LOUIS.
PA (UNIW) UNIV WASHINGTON.
XX Barenkamp SJ, St Geme JW;
PI WPI; 1996-455364/45.
DR N-PSDB; AAT41476.
XX Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in
PT vaccines against H. influenzae infection.
XX Claim 5; Page 66-73; 120pp; English.
XX Haemophilus adhesion protein HA2 (AAR99393) is associated with the
CC formation of surface fibrils involved in adhesion to various host
CC cells; it is also referred to hsf (Haemophilus surface fibrils).
CC Its amino acid sequence was deduced from a genomic DNA clone
CC (AAT41476) derived from Haemophilus influenzae type b strain C65.
CC Large quantities of recombinant HA2 can be produced in transformed
CC prokaryotic or eukaryotic host cells, for use in vaccines against
CC H. influenzae infection.
XX Sequence 2353 AA;
SQ

Query Match 4.9%; Score 138.5; DB 17; Length 2353;
Best Local Similarity 22.0%; Pred. No. 0.091;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;
QY 6 LGNNVGNHLIPPAPLPSTQDGAARGGTHLSISGTALGSRSLFSPLRNSMADSVDSRD 65
Db 1163 lggtag-----ndgtvinkdg-----litiangaagtdaen 1198
QY 66 IGPLTPNPSRLAAATSETCLLGGFEVLHDKGPL-----DILNTQIGSPAFRVEVQADGTHA 121
Db 1199 -----gntisvtkdgisagnkeitnvsalktykdtqnt-----adetqdkfha 1243
QY 122 AIGKNGLEVSVTLSPQEWSSLSQSIDTEGKNRFVFTGGRGSGHPMTVVASDIAEAR--- 178

Db 1244 av--knanevefv-----gkngatvsaktndngkhtvti--dvaekvgd 1284
QY 179 -----TRILAKLPDNGHGRQPKVDTRSVGVSAGIDDDGVVSETHST-TNSSVR--- 229
Db 1285 glektdgkiklkvdntdgnlltvd-tkgasvakgefnavtttdattaggtanergkv 1343
QY 230 -----SDPKFWVSVAIAAGLAGLAATGTAQALALTPPEPDPPTTDPDQAANA 277
Db 1344 vvksgngatatetdkkvtatvgdvakaind-aatfv-----kvenddsatiddsptddg 1396
QY 278 AESATK--DQLTQFAFKPNENKVNI-----ANGNAIPSGELXDDIVEQIAQAQKE 327
Db 1397 andaikagdtlilkagk---nlkvrdgknitfalandslvsksatvsdkl-----s 1444
QY 328 AGEVARQOAVESNAQAQRYEQHARRQBELQLSSGIGYSSALIVAGGIGAVTTALH 387
Db 1445 lgtngknvitsdtkglnfakdsktgddanihl-ngiastltdl-----lsgattnl- 1497
QY 388 RRNPAPAEQTTTTTHTVVOOQTGGIPQHKVALMPQERRRFSRDRDSQGSVASTHWSDS 447
Db 1498 -----gg-----ngitdnekkraasvkd----- 1515
QY 448 EYVNPYAEVGGARNSLSAHOPEHIYDEVAADPGYSVIONFSGSPVTRGLRGTPOGQIG 507
Db 1516 -vlnagvnrvvkpa-sannqveni-dfvt---ydtvdfvsgdkdtstvtveskdngkr 1569
QY 508 STYALLANGSGLRLMGGLTSGGE 531
Db 1570 tevkigaktsvikdhngkiftgke 1593

RESULT 9
AAB46770
ID AAB46770 standard; Protein; 3096 AA.
XX AAB46770;
XX 19-APR-2001 (first entry)
DT FANCIP3 protein fragment SEQ ID NO 10.
DE FANCIP3
KW FANCIP2; FANCIP3; antitumor; signal transduction cascade; immunogen;
KW Fanconi anemia complementation group A protein; DNA-repair defect;
KW cell-cycle disorder; cytopenia; tumorigenesis; tumor progression;
KW gene therapy.
XX Unidentified.
XX WO200100822-A2.
XX 04-JAN-2001.
XX 26-JUN-2000; 2000WO-EP05878.
XX 29-JUN-1999; 99DE-1029887.
XX (MULT-) MULTIGENE BIOTECH GMBH.
XX Gross HJ, Reuter T, Hanenberg H, Herterich S, Wagner M;
PI WPI; 2001-102892/11.
XX New nucleic acid encoding Fanconi anemia protein interacting proteins,
PT useful e.g. for diagnosis and treatment of DNA-repair defects -
XX Claim 1; Page 49-58; 60pp; German.
XX This invention describes a novel nucleic acid (I) which has antitumor
CC activity. (I) encode proteins that interact with the Fanconi anemia
CC protein of complementation group A, so may be part of the complex or
CC signal transduction cascade which causes the defect responsible for
CC Fanconi anemia. Polypeptides (II) encoded by (I) are used as immunogens

CC for preparation of antibodies (AB), also for identifying specific
 CC modulators (A). Any of (I) (or vectors or cells that contain (I)), (II),
 CC Ab and (A) are useful for: (a) diagnosis of diseases (or predisposition
 CC to them) associated with DNA-repair defects, cell-cycle disorders,
 CC cytopenia, tumorigenesis and tumor progression; and (b) for treatment
 CC (including by gene therapy) or prevention of these diseases.
 XX
 XX Sequence 3096 AA;

Query Match 4.9%; Score 136.5; DB 22; Length 3096;
 Best Local Similarity 21.0%; Pred. No. 0.19;
 Matches 144; Conservative 78; Mismatches 245; Indels 219; Gaps 33;
 QY 18 PAPPLPSQTDGARGTCHLISSTGALGSRSLFSLRNSMADSVSRDIP--GLPTNPSR 75
 Db 1916 pasplys---gialagl-----nlelysl-nglaslnysmetpheglnasple 1960
 QY 76 LAAATSETC-----LLG-----GPEVLHDKGP--LDLNTQ-----IG 106
 Db 1961 leglyltygliserleglylgyhisargargvallysaspalaleasnthrthrgla 2020
 QY 107 PSAFRVQADGTHAAIGKNGLEV--SVTLSPQWSSLSQSDT-----E 149
 Db 2021 snlyseleaspalatyrthrglnileaseraspelysargmetileaserlylelala 2080
 QY 150 GKNRFVFTGGR-----GGSGHPWTVASDIARFIRILAKLPDNDHG-----G 192
 Db 2081 glnvallysglnvalghisglnasmetileaserlearghisasnerargilehsvala 2140
 QY 193 ROPKVDTRSVGSGASGIDGWSETH--TSTNSSVRSRDPKFWVSGAAGLAGLA 249
 Db 2141 rpr--serargalaa-----nthrlealathrserasp---valserargly 2187
 QY 250 AT-----GIAQALALT-----PEPD-----PTTDPDQANAAE 279
 Db 2188 strplelleprglylagityrserilephethrglylnrleaspt----hrglnasp 2243
 QY 280 SATKQOLQEAQFNENQKVNIDANGNAIPSGELXDDIVEQIAQAKAGEVARQOAVES 339
 Db 2244 s--erasnvalaspaenglnle--glgthrcysserleglyhisargserprlegly 2298
 QY 340 NAAQQRVE-----DQHARRQBELQSSGIGYGLSALIVAGGIGAGVTTALH 387
 Db 2299 saspserprglyseraserthrerlelellelylsglnargl-----th 2351
 QY 388 RNQPAEQTTTTHTVVQQQG-----GIFQHKVALMPQERRRFS 428
 Db 2352 r-----serasphrilemetargalalelylsglleaspgiglylysi-lephelysas 2406
 QY 429 DRDS-----QGSVASTHWSDSSEVW---NPYAEVGGARNLSAHQPEE 470
 Db 2407 ntrpglythrlnthrglylsglasphthrserasnlleasn-----rargglnthrgit 2461
 QY 471 HIYDEVAADPGYSVQNFSGSPVTRGLIGTPGQGIQSTYALLANGSGRLMGGL----- 526
 Db 2462 hrservalasnalaserargserprgll-----yscysalaglnl-argglnlys 2512
 QY 527 -----TSGETAVSSVNAAPTGPVR 547
 Db 2513 argleasnseralaserargser 2538

RESULT 10
 ID AAW19629
 AC AAW19629 standard; Protein; 4630 AA.
 XX
 AC AAW19629;
 XX
 DT 01-SEP-1997 (first entry)
 XX
 DE Streptomyces venezuelae polyketide synthase.

KW Polyketide synthase; polyhydroxyalkanoate monomer synthase;
 KW polyhydroxybutyrate; biodegradable polymer; vep gene;
 KW metabolic engineering.
 XX Streptomyces venezuelae.
 XX WO97272711-A1.
 XX 26-JUN-1997.
 XX 18-DEC-1996; 96WO-US20119.
 XX 19-DEC-1995; 95US-0008847.
 XX (MINU) UNIV MINNESOTA.
 PA Sherman DH, Williams MD, Xue Y;
 PI WPI; 1997-341701/31.
 DR N-PSDB; AAT68715.
 DR Expression cassettes for production of polyhydroxyalkanoate(s) -
 PT provide wide range of biodegradable polymers for medical or
 PT industrial use
 XX Claim 55; Fig 23; 91pp; English.
 XX 3 Polypeptide sequences (AAW19629-30 and AAW00918) can be deduced from
 CC the vep ORF1 polyketide synthase (PKS) gene cluster (AAT68715) of
 CC Streptomyces venezuelae. The sequence data indicate that the PKS
 CC gene cluster encodes a polyene of 12 carbons. The vep gene cluster
 CC contains 5 PKS modules, plus a 5' loading module and a 3' end
 CC domain. Each of the sequenced modules includes a keto-ACP, an
 CC acyltransferase, a dehydratase, a keto-reductase and an acyl carrier
 CC protein domain. A novel expression cassette encoding the first
 CC module from the vep gene cluster and module 7 from the Streptomyces
 CC tyLP gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
 CC activity and can be used for PHA prodn. in host (esp. insect) cells
 CC for use as a biodegradable polymer.
 XX
 SQ Sequence 4630 AA;

Query Match 4.9%; Score 136.5; DB 18; Length 4630;
 Best Local Similarity 22.2%; Pred. No. 0.33;
 Matches 125; Conservative 49; Mismatches 203; Indels 187; Gaps 27;
 QY 23 PSQTDGAARGGTCHLISST--GALGSRSLFSLRNSMADSV--SRDIPGLPTNPSRLA 77
 Db 1156 prmhg--aegveghlltgttspvmsgrlayqilgtgavtdtaccssivalhavrslr 1214
 QY 78 AATSETCLLGFEVLHDKGPLDLINTQIGPSA-----FRVEVQADGTHAAIGKNGLEVSV 133
 Db 1215 qgesslallagatvmstpgmfefsrqrlaagrskafdsadgtsaegv--gllvve 1272
 QY 134 TLPSPQWSSLSQSDITEGKNRFVTGGRGSGHPMTVASDIA-----EAR 178
 Db 1273 rlsdae-----rnhpvlavirgsavngdngitapngpsq 1310
 QY 179 TRILAKLPDNDHGRQPKDVD--TRSVGVSASGIDGVVSETHSTTN-----S 226
 Db 1311 qrvirqalad--agltpadvdaveahgtgtrig--dpieaeailgtgydrgegaplqig 1366
 QY 227 SVRSDPKFWVSGATAAGLAGLAATGIAQALALTPEP-----DDPTTDDPOAANAASATK 283
 Db 1367 slksn-----iahaqaaagvgglikmvlamrhgvlprtlhvdprptr--vdweaggvellte 1421
 QY 284 DQLTQEAQFNENQKVNIDANGNAIPSGELXDDIVEQIAQAKAGEVARQOAVESNAQA 343
 Db 1422 er-----ewpetgr-----prraaisffisgtnahivveqapeage----- 1458
 QY 344 QORYEDQHARRQBELQSSGIGYGLSALIVAGGIGVTTALHRRNQPAE--QTTTTTT 401

Db 1459 -----aavttapeageageadtattt 1482
QY 402 HTVVQQTGGIPIQ-----HKVALMPQ--ERRRSDRDSOGSVASTHWSDSSE 448
Db 1483 paav-----gvpeprapvvsardaalarraqavrirtfdgr----- 1520
QY 449 VWPYARVGGARNSLAHQPEEH-----IYDEVAADPGYSVIONFSGSG-PVTGRLIG 500
Db 1521 ---pdvctvadglsraafefkhaaltatrdella--gldal-----grgeatqlvtg 1571
QY 501 TPGQGIQSTYALLANSGLRLGMG 524
Db 1572 eparaartaf-lfeggaqrvang 1594

RESULT 11
AAB40945
ID AAB40945 standard; Protein; 1532 AA.

AC AAB40945;
XX
DT 08-FEB-2001 (first entry)
XX Human ORFX ORF709 polypeptide sequence SEQ ID NO:1418.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

XX Homo sapiens.
OS
XX WO200058473-A2.
PN
XX 05-OCT-2000.
PD
XX 31-MAR-2000; 2000WO-US08621.
PF
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX (CURA-) CURAGEN CORP.

PA Shinkets RA, Leach M;
XX
PI WPI: 2000-602362/57.
XX N-PSDB; AAC75154.
DR
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 1203-1206; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;

CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX
SQ Sequence 1532 AA;

Query Match 4.8%; Score 135; DB 21; Length 1532;
Best Local Similarity 21.1%; Pred. No. 0.095;
Matches 127; Conservative 61; Mismatches 237; Indels 178; Gaps 28;

QY 22 LPSQTDGAA-RGCTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTN----- 72
Db 510 lpstsgaatrlvtgn--pstgaagt-----iprvpskvsaiqe---pgeptysshsttl 560
QY 73 PSRLAA-----TSEVCLLGGFEVLHDKGPLDI---LNTQIGPSAF-----RVEVQ 115
Db 561 pkttgagaqtwlqetgtg--eallisspsysvtqmkikatspspspmlrhtsqgitta 618
QY 116 ADGTHAAGEKNGLEVSVTLSPOEWSL-----QSIDTEGKNRFVFT 157
Db 619 pstnhslih-----stscpsqesavsrgrhtqapqtqesqtrsvspmdtkvtv 671
QY 158 GGRG--GSGH-----PMVTVASDIAEARTILAKILDPNHGGKQKDV-----DT 200
Db 672 pgsstasghspseivpqdaptisaattfapapt-----gdgthtqaptalqatpss 724
QY 201 RSVGVGSASGIDGVSFTHSTNTSSVSRDP-----KFWVSVGAIAAGLAGLAATGIAQA 256
Db 725 hdatlpgsggts-----lsktgaltlansvstpggpegqwsasa----- 765
QY 257 LALTPPEPDDPTTDPDQAA-----NAAESATKDOLTQEFKNPENQKVNIDANGNAIPSG 311
Db 766 -----stspdtaaaamththqaesteasgqtqts--epassgsrtrtsagatpss 812
QY 312 ELXDDIVEQIAQAKERAGEVARQAQVESNAQAQRYEDQHARRQEEQLQSSGICYGLSSA 371
Db 813 sgasgtlpsgsegistsgettr---fssnps-----rdshttgattellisasahg--a 861
QY 372 LIVAGGICAGVTTALHRRNQPAEOTTTTTTTHVVQQQTGGIPQHKVAL--MPQRRRFS 429
Db 862 ipvatgmassilvpgtfhptlseastagrtqgspspsasqetaalsrmaqtrts 921
QY 430 RRDQGSVASTHWSDSSEVVNPNYAIEVGGARNSLSAQHPEEHYDEVAADPGYSVIONFS 489
Db 922 rgsdtislas-qatdtfstv--pptpsitssgltsptqth----- 960
QY 490 GSGPVTGRLICTPGQGIQSTYALLANSGLRLGMGLTSGGET-----AVSSVNA 540
Db 961 -----tlpsgsggkttallisanatplpvttyassastghthplhvtassvstghat 1012
QY 541 PTP 543
Db 1013 plp 1015

RESULT 12
AAY27230
ID AAY27230 standard; Protein; 1978 AA.
XX
AC AAY27230;
XX

| | |
|--|---|
| Matches 123; Conservative 64; Mismatches 222; Indels 155; Gaps 27; | |
| Qy | 26 TDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPRLAATSETCL 85 |
| Db | 230 tdnagvnlftaypsgvgstl-----slttgtdt--ltgtanndttfvagevagaat 280 |
| Qy | 86 LGGFEVLHDKGPLDILN-----TOIGPSAFRVEVQADGTHAAIGEKNGLEVSVT 134 |
| Db | 281 ltvgtlsggagtdvlnwqaaavaltptgvtlsgietmmvtsgaaitntssgvtglta 340 |
| Qy | 135 LSPQEWSSLSQSIDT-EGKNRFVFTGGRGS-----GHPMVTVASDIAEARTRILAKLDPD 188 |
| Db | 341 lntntsgaqtvtagagqnlattataqaannvavdgranvtvas-----tgvt----- 389 |
| Qy | 189 NHGGRQPKVDTRSVGVGSAGIDGTVVSETHSTNTSSVRSRDPKFWVSGAIAAGLAGL 248 |
| Db | 390 -----gtttvgansaa---sgtvs---svansst-----tttgaia--vtgg 424 |
| Qy | 249 AATGIAQALALTPPEDDPTTTDPQAAANAASATKDOLTQEAFAKNPENQKVNIDANGNAI 308 |
| Db | 425 tadvagtag---navntltlqadvvtgnssttadvttqta-----aatagat 470 |
| Qy | 309 PSGELXD--DIVEIAQQAQKEAGEVAR-----QAVESNAQAQRYEDQHARRQBELQ 359 |
| Db | 471 vagrvngavtldsaasattagkiatvtlgsfgaatidssalttnl-----t 604 |
| Qy | 476 VAADPGYSVIONFS---GSGPVTGRIGTPOGQIQS-----TYALLANSQ-----GL 519 |
| Db | 605 aaltgltvtnsvgtlgaelatg-lvftggagrdsillgattkaimvgagddttvtssa 663 |
| Qy | 520 RLGMGLTSGG---ETAVSSVNA 540 |
| Db | 664 tlgaggsvnggdgtvlnvanvngs 687 |
| RESULT 14 | |
| AA44757 | |
| ID | AA44757 standard; Protein; 1026 AA. |
| AC | AA44757; |
| DT | 04-MAY-2000 (first entry) |
| DE | Caulobacter crescentus surface layer protein. |
| XX | Surface layer protein; S-layer secretion signal; antibiotic; vaccine; |
| KW | recombinant fusion protein cleavage; enzyme; protein polymer; |
| KW | antibacterial enzyme; foodstuff. |
| XX | Caulobacter crescentus. |
| XX | Key Location/Qualifiers |
| FT | Cleavage-site 692..693 |
| FT | /note= "Asp-Pro dipeptide present in S-layer secretion |
| FT | signal sequence. It is a site where a fusion |
| FT | protein comprising a target protein and the secretion |
| FT | signal is cleaved" |
| XX | WO200004170-A1. |
| PN | |
| XX | |
| PD | 27-JAN-2000. |
| XX | |
| PF | 14-JUL-1999; 99WO-CA00637. |
| XX | |
| PR | 14-JUL-1998; 98CA-2237704. |

| | |
|--|---|
| (UYBR-) UNIV BRITISH COLUMBIA. | |
| XX | Smit J; |
| XX | WPI: 2000-182434/16. |
| DR | N-PSDB; AA250079. |
| XX | Cleavage of Caulobacter produced recombinant fusion proteins useful for |
| PT | producing vaccine peptides |
| XX | Example 1; Pages 21-23; 33pp; English. |
| XX | The patent discloses a method for cleaving a recombinant fusion protein |
| CC | which is produced by Caulobacter and consists of Caulobacter surface |
| CC | layer (S-layer) protein (containing the C-terminal secretion signal) and |
| CC | a target protein heterologous to Caulobacter. The cleavage of target |
| CC | protein from the S-layer protein is carried out under mild acid |
| CC | conditions so that cleavage occurs at aspartate- |
| CC | proline dipeptide site without solubilising the protein. |
| CC | An cleavage is accomplished while the fusion protein is in |
| CC | an insoluble aggregate form which facilitates purification of the |
| CC | protein. The method is useful for producing pure proteins including |
| CC | recombinant human and animal therapeutic antibiotic and vaccine peptides, |
| CC | enzymes, protein polymers, and antibacterial enzymes for foodstuffs. |
| CC | The present sequence is a S-layer protein from C. crescentus. |
| CC | The S-layer secretion signal, corresponding to the C-terminal portion of |
| CC | the protein from amino acid 690 onwards, is fused with a target sequence |
| CC | for construction of a recombinant fusion construct which is expressed in |
| CC | Caulobacter. |
| XX | Sequence 1026 AA; |
| SQ | |
| Query Match 4.7%; Score 132.5; DB 21; Length 1026; | |
| Best Local Similarity 21.8%; Pred. No. 0.086; | |
| Matches 123; Conservative 64; Mismatches 222; Indels 155; Gaps 27; | |
| Qy | 26 TDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTNPRLAATSETCL 85 |
| Db | 230 tdnagvnlftaypsgvgstl-----slttgtdt--ltgtanndttfvagevagaat 280 |
| Qy | 86 LGGFEVLHDKGPLDILN-----TOIGPSAFRVEVQADGTHAAIGEKNGLEVSVT 134 |
| Db | 281 ltvgtlsggagtdvlnwqaaavaltptgvtlsgietmmvtsgaaitntssgvtglta 340 |
| Qy | 135 LSPQEWSSLSQSIDT-EGKNRFVFTGGRGS-----GHPMVTVASDIAEARTRILAKLDPD 188 |
| Db | 341 lntntsgaqtvtagagqnlattataqaannvavdgranvtvas-----tgvt----- 389 |
| Qy | 189 NHGGRQPKVDTRSVGVGSAGIDGTVVSETHSTNTSSVRSRDPKFWVSGAIAAGLAGL 248 |
| Db | 390 -----gtttvgansaa---sgtvs---svansst-----tttgaia--vtgg 424 |
| Qy | 249 AATGIAQALALTPPEDDPTTTDPQAAANAASATKDOLTQEAFAKNPENQKVNIDANGNAI 308 |
| Db | 425 tadvagtag---navntltlqadvvtgnssttadvttqta-----aatagat 470 |
| Qy | 309 PSGELXD--DIVEIAQQAQKEAGEVAR-----QAVESNAQAQRYEDQHARRQBELQ 359 |
| Db | 471 vagrvngavtldsaasattagkiatvtlgsfgaatidssalttnl-----t 518 |
| Qy | 360 LSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTHTVVQQQTGGIPOHKVAL 419 |
| Db | 519 --sgtgsi-----gigralta-----tptantltlnvnglt--ttgaidtseaa- 560 |
| Qy | 420 MPQERRRFSRRDSQGSVASTHWSDSSEVVNRYAEV-----GGARNSLSAQHPBEHIYDE 475 |
| Db | 561 -----addgfttiniagstaasiaslaadattinisdarvtitsh-----t 604 |
| Qy | 476 VAADPGYSVIONFS---GSGPVTGRIGTPOGQIQS-----TYALLANSQ-----GL 519 |
| Db | 605 aaltgltvtnsvgtlgaelatg-lvftggagrdsillgattkaimvgagddttvtssa 663 |

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:20:24 ; Search time 21.2 seconds
(without alignments)
533.212 Million cell updates/sec

Title: US-09-189-415A-2

Perfect score: 2800

Sequence: 1 MPICNLGNVNGNHLPPAP.....GETAVSSVNAAPTPGPVRFV 549

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 141 | 5.0 | 580 | 2 | US-08-672-564-1 |
| 2 | 138.5 | 4.9 | 2353 | 4 | US-09-377-155-33 |
| 3 | 138.5 | 4.9 | 2353 | 4 | US-08-913-942-4 |
| 4 | 134.5 | 4.8 | 1912 | 1 | US-08-409-995-4 |
| 5 | 134.5 | 4.8 | 1912 | 3 | US-08-685-467-4 |
| 6 | 133 | 4.8 | 2703 | 1 | US-08-185-432-19 |
| 7 | 132.5 | 4.7 | 1026 | 1 | US-08-194-230-7 |
| 8 | 132.5 | 4.7 | 1026 | 2 | US-08-614-377A-7 |
| 9 | 132.5 | 4.7 | 1026 | 4 | US-09-142-648B-7 |
| 10 | 128 | 4.6 | 1507 | 6 | 5268270-2 |
| 11 | 122.5 | 4.4 | 507 | 4 | US-08-860-635A-19 |
| 12 | 118.5 | 4.2 | 1600 | 2 | US-08-617-697-10 |
| 13 | 118 | 4.2 | 2441 | 1 | US-08-194-468-2 |
| 14 | 117.5 | 4.2 | 1536 | 1 | US-08-038-682-2 |
| 15 | 117.5 | 4.2 | 1536 | 1 | US-08-302-832-2 |
| 16 | 117.5 | 4.2 | 1536 | 2 | US-08-530-198-2 |
| 17 | 117.5 | 4.2 | 1536 | 2 | US-08-469-880-2 |
| 18 | 117.5 | 4.2 | 1536 | 2 | US-08-728-470-2 |
| 19 | 117.5 | 4.2 | 1536 | 2 | US-08-617-697-2 |
| 20 | 117.5 | 4.2 | 1536 | 4 | US-08-719-641-2 |
| 21 | 117.5 | 4.2 | 3816 | 4 | US-09-428-517-3 |
| 22 | 117 | 4.2 | 675 | 1 | US-08-386-495-10 |
| 23 | 117 | 4.2 | 675 | 5 | PCT-US96-02331-10 |
| 24 | 117 | 4.2 | 2441 | 3 | US-08-961-739-2 |
| 25 | 116 | 4.1 | 417 | 2 | US-08-672-564-9 |
| 26 | 116 | 4.1 | 2890 | 4 | US-09-413-814-67 |
| 27 | 116 | 4.1 | 3798 | 3 | US-09-335-409-6 |

| | | | | | | |
|----|-------|-----|------|---|-------------------|--------------------|
| 28 | 114.5 | 4.1 | 1186 | 2 | US-08-861-464-8 | Sequence 8, Appli |
| 29 | 114.5 | 4.1 | 1186 | 2 | US-08-396-001-8 | Sequence 8, Appli |
| 30 | 114.5 | 4.1 | 1186 | 4 | US-09-323-433A-8 | Sequence 8, Appli |
| 31 | 114.5 | 4.1 | 1529 | 2 | US-08-728-470-10 | Sequence 10, Appli |
| 32 | 114.5 | 4.1 | 1529 | 4 | US-08-719-641-10 | Sequence 10, Appli |
| 33 | 114.5 | 4.1 | 5588 | 4 | US-09-036-987A-6 | Sequence 6, Appli |
| 34 | 113 | 4.0 | 1004 | 4 | US-08-916-352-2 | Sequence 2, Appli |
| 35 | 113 | 4.0 | 1043 | 3 | US-08-928-361B-30 | Sequence 30, Appli |
| 36 | 113 | 4.0 | 1721 | 3 | US-08-700-651-5 | Sequence 5, Appli |
| 37 | 113 | 4.0 | 1721 | 3 | US-08-928-361B-6 | Sequence 6, Appli |
| 38 | 111.5 | 4.0 | 619 | 4 | US-09-066-046-2 | Sequence 2, Appli |
| 39 | 111.5 | 4.0 | 1794 | 6 | 5183745-6 | Patent No. 5183745 |
| 40 | 110.5 | 3.9 | 1848 | 4 | US-08-296-791-6 | Sequence 6, Appli |
| 41 | 110.5 | 3.9 | 1848 | 5 | PCT-US95-10661A-6 | Sequence 6, Appli |
| 42 | 110.5 | 3.9 | 2516 | 3 | US-08-374-077C-2 | Sequence 2, Appli |
| 43 | 110.5 | 3.9 | 2516 | 4 | US-08-895-590-2 | Sequence 2, Appli |
| 44 | 110 | 3.9 | 4551 | 3 | US-09-320-878-1 | Sequence 1, Appli |
| 45 | 109.5 | 3.9 | 475 | 1 | US-08-350-741-2 | Sequence 2, Appli |

ALIGNMENTS

RESULT 1
US-08-672-564-1
; Sequence 1, Application US/08672564
; Patent No. 5824503
; GENERAL INFORMATION:
; APPLICANT: KURUME, Yoko
; APPLICANT: IZU, Hiroyuki
; APPLICANT: IZUMI, Yoshiya
; APPLICANT: SANO, Mutsumi
; APPLICANT: KATO, Ikunoshin
; APPLICANT: ITO, Makoto
; TITLE OF INVENTION: GENE ENCODING ENDOGLYCERAMIDASE ACTIVATOR
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/672.564
; FILING DATE: 28 JUNE 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WEINER, Marc S.
; REGISTRATION NUMBER: 32,181
; REFERENCE/DOCKET NUMBER: 1422-0263p
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-672-564-1

Query Match 5.0%; Score 141; DB 2; Length 580;
Best Local Similarity 23.9%; Pred. No. 0.00043;
Matches 84; Conservative 45; Mismatches 126; Indels 96; Gaps 16;


```

;
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
US-08-409-995-4

Query Match      4.8%  Score 134.5; DB 1; Length 1912;
Best Local Similarity 22.0%; Pred. No. 0.01;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;

QY 6 LGNNVNGNHLPPAPPPLPSQDTGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRD 65
Db 1163 LGGTANG-----RNDGTGVINKDG-----LTITLANGAAAGTGDASN 1198
QY 66 IPGLPTNPSRLAAATSETCLLGGFEVLHDKGPL-----DILNTQIGPSAFRVEVQADGTHA 121
Db 1199 -----GNTISVTKDGISAGNKEITNVKSALTKYKDTQNT-----ADETQDKKEFHA 1243
QY 122 AIGKNGLEVSVTLSPQWSSLSQSDTEGKNRFVFTGGRGSGHPMTVVASDIAR-----178
Db 1244 AV--KNAEVEFV-----GKNATVSAKTNDNGKHVTI--DVAEAKVGD 1284
QY 179 -----TRILAKLDPDHGGRPKVDTRSVGVSAGIDGWSVETHTST--TNSSVR---229
Db 1285 GLEKDTGKIKLVNDTGNLLTVDA--TKGASVAKGEFNAVTTDATTAOGTNANERGV 1343
QY 230 -----SDPKFWVSVGATAAGLAGLAATGIAQALALTPEDDPTTDDPQAANA 277
Db 1344 VVKGSGATATETDKKVVATGVDVAKAIND--AATFV-----KVENDDSATIDDSPTDGG 1396
QY 278 AESATK--DQLTOEAFKNPENQKNID-----ANGNAIPSGELXDDIVEQIAQAQAKE 327
Db 1397 ANDALKAGDTTLTKAGK--NLKVKRDGKNITFALANDLSVKSATVSKL-----S 1444
QY 328 AGEVARQAQVESNAQAQORYEDOHARROEELQSLSSGIGYGLSSALIVAGGAGVTTALH 387
Db 1445 LGTNGKNVNIITSDTKGLNFAKDSKTGDDANIHL--NGIASTLTDTL-----LNSGATTNL- 1497
QY 388 RRNPQAEQTTTTTTHVVVQQQTGGIPQHKVALMPQERRRSDRRDSQGSVASTHWSDS 447
Db 1498 -----GG-----NGITDNEKKRAASVKD-----1515
QY 448 EVVNPYAEVGGARNSLSAHOPEEHIYDEVAADPGYSVIONFSGSGPVTRGLIGTPGQGIQ 507
Db 1516 -VLNAGWNVGVKPA-SANNOVENI-DFVAT---YDVTDFVSGDKDTTSVTVESKDNGKR 1569
QY 508 STYALLANSGLRLGMGLTSGGE 531
Db 1570 TEVKIGAKTSVIKDHNGKLTGKE 1593

RESULT 4
US-08-409-995-4
; Sequence 4, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989

```

```

;
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
US-08-409-995-4

Query Match      4.8%  Score 134.5; DB 1; Length 1912;
Best Local Similarity 22.0%; Pred. No. 0.01;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;

QY 6 LGNNVNGNHLPPAPPPLPSQDTGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRD 65
Db 1162 LGGTANG-----RNDGTGVINKDG-----LTITLANGAAAGTGDASN 1197
QY 66 IPGLPTNPSRLAAATSETCLLGGFEVLHDKGPL-----DILNTQIGPSAFRVEVQADGTHA 121
Db 1198 -----GNTISVTKDGISAGNKEITNVKSALTKYKDTQNT-----ADETQDKKEFHA 1242
QY 122 AIGKNGLEVSVTLSPQWSSLSQSDTEGKNRFVFTGGRGSGHPMTVVASDIAR-----178
Db 1243 AV--KNAEVEFV-----GKNATVSAKTNDNGKHVTI--DVAEAKVGD 1283
QY 179 -----TRILAKLDPDHGGRPKVDTRSVGVSAGIDGWSVETHTST--TNSSVR---229
Db 1284 GLEKDTGKIKLVNDTGNLLTVDA--TKGASVAKGEFNAVTTDATTAOGTNANERGV 1342
QY 230 -----SDPKFWVSVGATAAGLAGLAATGIAQALALTPEDDPTTDDPQAANA 277
Db 1343 VVKGSGATATETDKKVVATGVDVAKAIND--AATFV-----KVENDDSATIDDSPTDGG 1395
QY 278 AESATK--DQLTOEAFKNPENQKNID-----ANGNAIPSGELXDDIVEQIAQAQAKE 327
Db 1396 ANDALKAGDTTLTKAGK--NLKVKRDGKNITFALANDLSVKSATVSKL-----S 1443
QY 328 AGEVARQAQVESNAQAQORYEDOHARROEELQSLSSGIGYGLSSALIVAGGAGVTTALH 387
Db 1444 LGTNGKNVNIITSDTKGLNFAKDSKTGDDANIHL--NGIASTLTDTL-----LNSGATTNL- 1496
QY 388 RRNPQAEQTTTTTTHVVVQQQTGGIPQHKVALMPQERRRSDRRDSQGSVASTHWSDS 447
Db 1497 -----GG-----NGITDNEKKXAA-----SVK 1513
QY 448 EVVNPYAEVGGARNSLSAHOPEEHIYDEVAADPGYSVIONFSGSGPVTRGLIGTPGQGIQ 507
Db 1514 DVLNAGWNVGVKPA-SANNOVENI-DFVAT---YDVTDFVSGDKDTTSVTVESKDNGKR 1568
QY 508 STYALLANSGLRLGMGLTSGGE 531
Db 1569 TEVKIGAKTSVIKDHNGKLTGKE 1592

RESULT 5
US-08-409-995-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-685-467-4

Query Match 4.8%; Score 134.5; DB 3; Length 1912;
Best Local Similarity 22.0%; Pred. No. 0.01;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;

QY 6 LGNNVNGNHLIPAPPPLPSQTDGAARGTGHLSITGALGSRSLFSPLRNSMADSVSRD 65
DB 1162 LGGTANG-----RNDTGTIVKDG-----LTIITLANGAAAGTDA 1197
QY 66 IPGLTPNPSRLAAATSETCLLGGFEVLHDKGPL-----DILNTQIGPSAFRVEQADGTHA 121
DB 1198 -----GNTISVTKDGISAGNEITNVKSAKYKDTQNT-----ADETDQKEFHA 1242
QY 122 AIGERNGLVSVTLSPQEWSSLSQIDTEGKNRFVFTGGRGSGGHPMVTASDIABAR--- 178
DB 1243 AV--KNAEVEFV-----GKNKATVSAKTDNNGKHVTI--DVAAEKVGD 1283
QY 179 -----TRILAKIDPDNHNHGRQPKVDYTRSVGVSGASGIDGCVSWETHST--TNSSVR--- 229
DB 1284 GLEKTDGKIKLVKVDNTDGNLLTVDA--TKGASVAKGEFNATVDTAATQGTNANERGV 1342
QY 230 -----SDPKFVWSVGATAAGLAGLAATGIAQALALTPPEPDDPTTDDPQAA 277
DB 1343 VYKGSNGATATETDKKKVATGVDVAKAIND--AATFV-----KVENDDSATIDDSPTDDG 1395
QY 278 AESATK--DQLTQEAFAKPNQKVNID-----ANGNATPSGELXDDIVEQTAQQA 327
DB 1396 ANDALKAXDITLTKAGK--NLKVKRDGKNITFALANDLSVKASATVSKL-----S 1443
QY 328 AGEVARQAQAVESNAQOQRYEDOHARRQEEQLQSSGIGYGLSSALIVAGGICAGVTTALH 387
DB 1444 LGTNGKNVITSDTKGLNFAKDSKTDGDANTHL--NGIASTLTDIL-----LNSGATTNL- 1496
QY 388 RRNPQAEQTTTTTHTVVOOQTGGIPQHKVALMPOERRRDRDRSQSVASTHWSDDSS 447
DB 1497 -----GG-----NGITDNEKKXAA-----SVK 1513
QY 448 EVNYPYAEVGGARNLSAHOPEHIYDEVAADPGYSVIONFSGSGPVTGRLIGTPCGQIQ 507
DB 1514 DVLNAGWNVRGVKPA--SANNQVENI--DFVAT---YDVTDFVSGDKDTTSVTESKDNGR 1568
QY 508 STYALLANSGLRLGMGGITSGGE 531
DB 1569 TEVKGAKTSVTKDHNGKLFKGE 1592
```

```

RESULT 6
US-08-185-432-19
; Sequence 19, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Busseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2703 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-185-432-19
```

```

Query Match 4.8%; Score 133; DB 1; Length 2703;
Best Local Similarity 19.8%; Pred. No. 0.024;
Matches 119; Conservative 83; Mismatches 193; Indels 206; Gaps 28;

QY 4 GNLGNNVNGNHLIPAPPPLPSQT--DGAAR-----GGTGHLSITGALGSRSLFSPLRN 55
DB 2177 GNGGNNNGN-----ASGKQSNQAKQAQKAKLIEGSPDNGLDATGSLRRK---ASSKK 2229
QY 56 SMADSVSRDIPGLTPNPSRLAAATSETCLLGGFEVLHDKGPLDILATQIGPSAFRVEQ 115
DB 2230 TSAASKAANLNGL--NPGQLT-----GGVSGVPGVPP---TNSAVQAAAAAAAV 2275
QY 116 ADGTHAAIGERNGLVSVTL--SPOEWSLSQSIDTEGKNRFVFTGGRGSGGHPMVTASDI 174
DB 2276 AAMSHLEGGSPVGMGNLPSPYDTSSMY----- 2306
QY 175 AEARTILAKIDPDNHNHGRQPKVD-----TRSVGVSGASGIDGCVVSEHTSTNSVR 229
DB 2307 -NMAAAPLANGNP--NTGAKOPPSYEDCIKNAQSMQSLQGNGL--DMIKLDNVAYSMGSPFQ 2363
QY 230 SD-----PKFWVSVGAIAAGLAGL-----AATG----- 252
DB 2364 QELINGOGLGMNGNGQNRGPGVLPGLCGGLSGAGNGNSREQGLSPPSYNSQSPHS 2423
QY 253 IAQALALTP-----PDDPTTDDPQAAANAAESATKDQLTQEA 291
```


Thu Sep 27 15:29:41 2001

REGISTRATION NUMBER: 34053
REFERENCE/DOCKET NUMBER: 08106/002001
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1026 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-614-377A-7

Query Match 4.7%; Score 132.5; DB 2; Length 1026;
Best Local Similarity 21.8%; Pred. No. 0.0059;
Matches 123; Conservative 64; Mismatches 222; Indels 155; Gaps 27;

QY 26 TDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDRDIPGLTPNPSRLAAATSETCL 85
DB 230 TDNAAGVNLFTAYPSSGVSGSTL-----SLTGTGDT--LTGTANNDTFVAGEVAGAAAT 280

QY 86 LGGFVFLHDKGPLDILN-----TGIGPSAFRVEQADGTHAAIGKNGLEVSVT 134
DB 281 LTVGDTLSSGAGTDVNLNVQAAAVTALPTGVTISGIETMNVTSAGAITLNTSSGVTGLTA 340

QY 135 LSPQEWSSLSQIDT--EGKNRFVFTGGRGGS-----GHPMVTVASDIAEARTILAKLDPD 188
DB 341 LNTNTSGAAQTAVAGAGNLATTTAAQAANNVAVDGRANVTAS-----TGVTST----- 389

QY 189 NHGGROPKDVDRSVGSGASGIDGVVSETHSTNTSSSVSRDPKFWVSGAIAAGLAGL 248
DB 390 -----GTTTVGANSAA---SGTVS---VSVANSST-----TTTGAIA--VTGG 424

QY 249 AATGIAQALALTPEDPDDTTTDPQAAANAESATKDQLTQEAFAFNKPNQKYNIDANGNAI 308
DB 425 TAVTVAQTAG---NAVNTLTQADVTVTGNSSTTAVTTQTA-----AATAGAT 470

QY 309 PSGLXD--DIVEQIAQAQAKEAGEVAR-----QOAVESNAQAQORYEDQHARRQEELQ 359
DB 471 VAGRVNGAVTITDSAAASATAGKIATVTLGSFGAATIDSSALTIVNL----- 518

QY 360 LSSGIGYGLSSALIVAGGIGACVTTALHRRNQAPAEQTTTTTHTVVQQOQGGIPQHKVAL 419
DB 519 --SGTGTSL-----GIGRGALTA-----TPTANTLTNLVNGLT--TTGAITDSEAA- 560

QY 420 MPQERRFRDRDSQGSVASTHWSDSSEVVNPAEY---GGARNSLSAQHPPEHIYDE 475
DB 561 -----ADGFTTINIAGSTASTIASLVAADATTLNISGDARVTISH-----T 604

QY 476 VAADPGYSVIONFS---GSGPVTGRIGTPGQGIQS-----TYALLANSGL-----GL 519
DB 605 AALTGITVNSVGTALGAEALATG-LVFTGGAGRDSILLGATTKAIVMGAGDDTTVTVSSA 663

QY 520 RLGMGGLTSGG---ETAVSSVNA 540
DB 664 TLGAGGSVNGGDDTDLVANVNGS 687

RESULT 9
US-09-142-648B-7
Sequence 7, Application US/09142648B
Patent No. 6210948
GENERAL INFORMATION:
APPLICANT: Smit, John
APPLICANT: Bingle, Wade H.
APPLICANT: No. 6210948ellini, John F.
TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULOPHYTES
FILE REFERENCE: 08106/002002
CURRENT APPLICATION NUMBER: US/09/142,648B
CURRENT FILING DATE: 1999-03-30
PRIOR APPLICATION NUMBER: PCT/CA97/00167

PRIOR FILING DATE: 1997-03-10
PRIOR APPLICATION NUMBER: US 07/614,377
PRIOR FILING DATE: 1996-03-12
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 1026
TYPE: PRT
ORGANISM: Caulobacter crescentus
US-09-142-648B-7

Query Match 4.7%; Score 132.5; DB 4; Length 1026;
Best Local Similarity 21.8%; Pred. No. 0.0059;
Matches 123; Conservative 64; Mismatches 222; Indels 155; Gaps 27;

QY 26 TDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDRDIPGLTPNPSRLAAATSETCL 85
DB 230 TDNAAGVNLFTAYPSSGVSGSTL-----SLTGTGDT--LTGTANNDTFVAGEVAGAAAT 280

QY 86 LGGFVFLHDKGPLDILN-----TGIGPSAFRVEQADGTHAAIGKNGLEVSVT 134
DB 281 LTVGDTLSSGAGTDVNLNVQAAAVTALPTGVTISGIETMNVTSAGAITLNTSSGVTGLTA 340

QY 135 LSPQEWSSLSQIDT--EGKNRFVFTGGRGGS-----GHPMVTVASDIAEARTILAKLDPD 188
DB 341 LNTNTSGAAQTAVAGAGNLATTTAAQAANNVAVDGRANVTAS-----TGVTST----- 389

QY 189 NHGGROPKDVDRSVGSGASGIDGVVSETHSTNTSSSVSRDPKFWVSGAIAAGLAGL 248
DB 390 -----GTTTVGANSAA---SGTVS---VSVANSST-----TTTGAIA--VTGG 424

QY 249 AATGIAQALALTPEDPDDTTTDPQAAANAESATKDQLTQEAFAFNKPNQKYNIDANGNAI 308
DB 425 TAVTVAQTAG---NAVNTLTQADVTVTGNSSTTAVTTQTA-----AATAGAT 470

QY 309 PSGLXD--DIVEQIAQAQAKEAGEVAR-----QOAVESNAQAQORYEDQHARRQEELQ 359
DB 471 VAGRVNGAVTITDSAAASATAGKIATVTLGSFGAATIDSSALTIVNL----- 518

QY 360 LSSGIGYGLSSALIVAGGIGACVTTALHRRNQAPAEQTTTTTHTVVQQOQGGIPQHKVAL 419
DB 519 --SGTGTSL-----GIGRGALTA-----TPTANTLTNLVNGLT--TTGAITDSEAA- 560

QY 420 MPQERRFRDRDSQGSVASTHWSDSSEVVNPAEY---GGARNSLSAQHPPEHIYDE 475
DB 561 -----ADGFTTINIAGSTASTIASLVAADATTLNISGDARVTISH-----T 604

QY 476 VAADPGYSVIONFS---GSGPVTGRIGTPGQGIQS-----TYALLANSGL-----GL 519
DB 605 AALTGITVNSVGTALGAEALATG-LVFTGGAGRDSILLGATTKAIVMGAGDDTTVTVSSA 663

QY 520 RLGMGGLTSGG---ETAVSSVNA 540
DB 664 TLGAGGSVNGGDDTDLVANVNGS 687

RESULT 10
5268270-2
Patent No. 5268270
APPLICANT: Meyer, Thomas F.; Halter, Roman; Pohlner, Johannes
TITLE OF INVENTION: PROCESS FOR PRODUCING PROTEINS USING GRAM-NEGATIVE HOST CELLS
NUMBER OF SEQUENCES: 6
CURRENT APPLICATION DATA: US/07/171,872
APPLICATION NUMBER: US/07/171,872
FILING DATE: 01-JUL-1987
SEQ ID NO: 2
LENGTH: 1507
5268270-2

Query Match 4.6%; Score 128; DB 6; Length 1507;

```
Best Local Similarity   21.4%; Pred. No. 0.027;
Matches 104; Conservative    60; Mismatches 180; Indels 142; Gaps  22;
```

QY 34 TGHLSISGTGALGSRLSPLRNSWADSV--SRDIPGLPTNPRLAAATSETCLLGGFEVL 92
 ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 757 TGYVTCNTGNLSKAL-----NFDXATKINGVNLNQNLAALVLGKAALMGKTQ-- 805

QY 93 HDKGPLDLINTQTIGPSAFRVEVDAGTHAAIGEKNGLEVSVTLSQPQMSSLSQSDITEGKN 152
 || | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 806 -----CQGNSRVSLNQHSKWHLTGDSOVHNLSDLASHIHLNNASDAQSANKY 852

QY 153 RFVFTGGRSGG--HPMVTVVASDIABEARTIRILAKLDPNHGGRPKDVDTRSVGVSASG 210
 : ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 853 HTIKINLSNGHGPHYETDLAKNLGD---KYLVR---EASAGHYQLHVQNKTT-GEPMOEG 905

QY 211 ID--DGVVSETHSTTTSNSSVRSDPKFVWSVGAIAAGLAGLAAT----GIAQAALALTPEP 263
 ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 906 LDLFDA-----SSVQDRSLFVSLANHHYVDLGALRYTIKTENGITRLNYPNAGN 954

QY 264 DDPHTTDPAANAASESATKDQLTQEA-FKNPENOKVINIDANGAIPSSELXDIVEQIA 322
 || | ||| ||| ||| ||| :|| :|| :|| :|| :|| :|| :||
Db 955 GRPVKPAPSPAANTASQAQRATQTDGAQIAKPNIV-----APPSPQANQ--AEELR 1005

QY 323 QOAKEAGEVARQQAVESNAQAQRYED-----QHARRQEBLQISSGIGYGLSSALIVAG 376
 |||| :|||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1006 QQAK-AEQVKRQQAEEKVARKQDEEAKRAABIARQOEER-----KAEELAAK 1055

QY 377 GIGAGVTTALHRRNQPAEOFTTTTTTHVVQOQTGGIQPHKVALMPQERR----- 426
 ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1056 QKAEBEKARELARQKAEAS-----HOANAKPRRRRRRAILLPPRPAP 1098

QY 427 -FS-----DRRDQSOGSVASTHWSDSS-----SEVVPYAIEVGGARNSSLASHOPE 469
 || || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1099 VFLDDDYDAKDN-----SESSIONLARVPIRMGRELINDEYEI-----PL 1138

QY 470 EHIYDE 475
 | :||
Db 1139 EELEDE 1144

RESULT 11
US-08-860-635A-19
 ; Sequence 19, Application US/08860635A
 ; Patent No. 6143878
 ; GENERAL INFORMATION:
 ; APPLICANT: Koopman, Peter
 ; APPLICANT: Goodfellow, Peter
 ; TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
 ; TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Scully, Scott, Murphy & Presser
 ; STREET: 400 Garden City Plaza
 ; CITY: Garden City
 ; STATE: NY
 ; COUNTRY: U.S.A.
 ; ZIP: 11530
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq Version 1.5
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/860,635A
 ; FILING DATE: 29-MAY-1997
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: AU PM9714
 ; FILING DATE: 29-NOV-1994
 ; APPLICATION NUMBER: AU PM9835
 ; FILING DATE: 05-DEC-1994
 ; APPLICATION NUMBER: PCT/AU95/00799
 ; FILING DATE: 29-NOV-1995

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/617,697
;; FILING DATE: 01-APR-1996
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/302,832
;; FILING DATE: 05-OCT-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US PCT/US93/02166
;; FILING DATE: 16-MAR-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Berkstresser, Jerry W
;; REGISTRATION NUMBER: 22,651
;; REFERENCE/DOCKET NUMBER: 1038-557
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 415-0810
;; TELEFAX: (703) 415-0813
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1600 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-617-697-10

Query Match 4.2%; Score 118.5; DB 2; Length 1600;
Best Local Similarity 19.4%; Pred. No. 0.2; Indels 157; Gaps 26;
Matches 124; Conservative 84; Mismatches 275;
QY 1 MPIGNL--GN-----NVNGNHLPPAPLPSPQDGAARGGTGHLISSTGALGSRSL 49
Db 878 LFLNLISVEGNLSLGNANIVGN-----LSTAEDSTKFEASDNLNITGFTNNGT 929
QY 50 FSLRLNSMADSVDSRIDPELPNPNRLAAATSETCLLGGFEVLHDKGPDILN-----T 103
Db 930 ANINIKGVVVLGDIINNGKL--NITNASGTQKTIING--NITNEKGLDNKINKADAEI 985
QY 104 QIGPSAFRVE---VOAGDTHAA--IGEKNGLE-----VSVTLSPQEWSSLSQSI 146
Db 986 QIGGNISQKEGNLTSSDKVNTNQTITKAGVEGGRSDSSSEANANLIQIKELAGDL 1045
QY 147 DTGKRNRFVFTGGGGGHPMVTVASDIAEARTRIL-----AKLDPDNHGGQPKDVTR 201
Db 1046 NISGFNKAEIT-AKNGSDLTICNAGSNADAKKVTDFKDKSKISTDGHNVTLNSEVKT 1104
QY 202 SVVGVGSASIDGCVYSEHTSTNSVSRDPKFWVSVA---IAAGLAGLAATGIAQAL 257
Db 1105 NGSSNAGNDNSTGLTISAKDVTNNVNNVTSKHTINISAAAGNVTTKEGTINATTSVEVT 1164
QY 258 ALTPEDPDPTTDPQAAANAESATKDQTOE-AFKNPENKQVNI-----D 302
Db 1165 A-----QNGTIKGNITSQNVTVTATENLVTTENAVINATSGTVNISTKTDIKGGIEST 1219
QY 303 ANGNATPSGEL--XDDIVQIAQAKEAGEVARQ-----QAVESNAQAQRYEDOHARRQ 355
Db 1220 GNVNITASGNLTKVSNITGQDVTVTADAGALATTAGSTISATTTGNANITTKTGDKNG-- 1277
QY 356 EELQLSSG-IGYCLLSALIVAGGIGAG---VTTALHRRNQPAEQTTTTTHFVWQQTGG 411
Db 1278 --VSSSGSVTLVATGATLAVGNISNTVTTADSGKLTSTVGSTINGNSTVTSQSG- 1334
QY 412 IPQHKVALMPQERRRFSRRDSQGSVASTHWSDDSEVVNPNYAIEVGARNLSAHOPEH 471
Db 1335 -----DIEGTI-----SGNTVNTASTGDLTIGNSAKVEAKN 1366
QY 472 IYDEVAADPGYSVIQNFSG-----SGPV-----TGLRIGTPGQG 505
Db 1367 GAATLTAESGKLTOTGSSITSSNGOTTITAKDSSSTAGNINANVTLNTGTITTTGDSK 1426
QY 506 IQSTYALLANGGLRLG-----MGGLTSGGETAVSSVNA 540
Db 1427 INAT-----SGTLTINAKDAKLDAASGDRTVVVNATNAS 1460

RESULT 13
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
; TITLE OF INVENTION: RESPONSIVE GENES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,468
; FILING DATE: 10-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9672
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)-546-4737
; TELEFAX: (619)-546-9392
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-194-468-2

Query Match 4.2%; Score 118; DB 1; Length 2441;
Best Local Similarity 19.0%; Pred. No. 0.43;
Matches 101; Conservative 70; Mismatches 177; Indels 184; Gaps 26;
QY 1 MPIGNLGNVNGNHLPPAPLPSPQDGAARGGTGHLISSTGALGSRSLFSLRLNSMADS 60
Db 45 IPNGEL-SLLNSGNLYPDAAASKHQLSELLRGGSGSSINP--GIGNVSASSPVQOGLGGQ 101
QY 61 VDSR-----DIPGL-----PTNPSRLA----- 77
Db 102 AOGQPNSTMASLGAMGKSPLNQGDSTPNLPKQAASTSGPTPPASQALNPQAOKQVGLV 161
QY 78 ---AATST---CLLGGFEVLHDKGPDILNTQIGPSAFRVEVQA-----DGHAAIGE 125
Db 162 TSSPATSTQGTGICMNFNQTHTP---GLLNSNSGHSMLNMQAQQOQAVVNGSLGAGR 217
QY 126 KNG-----LEVSVTLSPQ---EWSLSQSIDTECKNRFVFTGGRGSGH 165
Db 218 GRGAGMPYPAPAMQCATSSVLAETLTQVSPQAGHAGLNTAQAGGTMKGMTGTTSPFGQ 277
QY 166 PMVTVASDIAEARTRILAKLPDNHGGQPKDVTRSVGVGSASIDGCVYSETH----- 220
Db 278 PF-----SQTGGQGM-----GATGVNPQLASKOSMNSL 306
QY 221 ----TSTTNSSVSRDP---KFWVSVGAIAAGLAGLAATGIAQALALTPEDPTTDPDQ 273
Db 307 PAFPTDIDIKNTSVTTVPNNMSQLOTSVGIVP-----TQAIATGP-----TADPE- 348
QY 274 AANAASATKDQTOE-----AFKNPENKQVNIANDANGNAIPSGELXDDIVEQIAQ- 325

Db 349 -----KRLIQQLVLLHAKCQRREQANCEVRACSLPHCRTMKNVNLNHTHQA 399
 QY 326 KEAGEVAR--QOAVESNAQAQRYE-----DOHARROEEELQSSGIGYGLSSALIV 374
 Db 400 PKACQVAHCASSRQIISHWKNCTRHDPCVCLPLKNASDKRNQQTILGSPAS-GIQNTI-- 456
 QY 375 AGGIGAGVTTALHRRN-OPAEQTTTTTTHVVQQTGGIP---OHKVALMPQ 422
 Db 457 -GSVGAGQONATSLNPNPIDPSSMORAYAL-----GLPYNQDQLOLQFO 502

RESULT 14
 US-08-038-682-2
 ; Sequence 2, Application US/08038682
 ; Patent No. 5549897
 ; GENERAL INFORMATION:
 ; APPLICANT: BARENKAMP, STEPHEN J
 ; APPLICANT: ST. GEME III, JOSEPH W
 ; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
 ; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; STREET: Bldg. 1
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/038.682
 ; FILING DATE: 16-MAR-1993
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BERKSTRESSER, JERRY W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-293
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0813
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1536 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-038-682-2

Query Match 4.2%; Score 117.5; DB 1; Length 1536;
 Best Local Similarity 19.7%; Pred. No. 0.23; 239; Indels 213; Gaps 26;
 Matches 129; Conservative 73; Mismatches 239; Indels 213; Gaps 26;
 QY 37 LISSTGALGSRSLFPLRNSMADSVSDRDPGLPTNPRLAAATSETCLLGGF----- 89
 Db 223 VLSVNG--GSISLAGQKTTISDIIN-----PTTYSIAAPENAVNLGDIFAKGGNI 273
 QY 90 -----EVLHDKGPDLIDLTQIGPSAFRVEQADTHAAIGE----- 125
 Db 274 NVRAATIRNQKLSADSVSKDKSGNVLVLSAKEGEAEIGGVISAQNOQAKGKLMITGDKV 333
 QY 126 -----KNGLEV-----SVTLSPQE-----W 140
 Db 334 FLTKTGAVIDLSGKEGGTYLGDGERGKNGIQIAKKTSLKGTSTINVSKEKGRAIIV 393
 QY 141 SSLQSID----TEGKNRFVFTGG-RGGSGHPMVTVASDIAEARTRILAKLPDN----- 189

Db 394 GDIALIDGNIQAQSGDIAKGTGGFVETSGHDLFIKDNAIVDAKEWL-----LDFDNVSNAE 450
 QY 190 HGGQPKDQVTRSVGVG--SASGIDGCVVSETHSTTNSSVSRSDPKFW----- 235
 Db 451 TAGRSNTSEDEYTGSGNSASTPKRNKEKTLTNTLESILKKGTFVNTANQRIYVNS 510
 QY 236 --VSVGATAAGLAGLAATGIAQALALTPEDDPDTTDPDQAANAASATKDQLT----- 287
 Db 511 INLSNGSLTWSEGRSGGV-----EINNDITGDDTRCAN-----LTIYSGW 554
 QY 288 QEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQAQAKEAGEVARQOAVESNAQAQRY 347
 Db 555 VDVHKN-----ISLGAQONI-----NITAKODIAFEKGSNOVITGQ----- 591
 QY 348 EDQHARQOEELQSSGIGYGLSSALIIVAGGIGAGVTTALHRRNQPA-----EQTTT----- 398
 Db 592 -----TITSGNQKGFNNVSLNGTSGLGQFTTKRTNKYAITNKFEGLTNISGK 640
 QY 399 TTTHTVVQQTGGIPOHK-----VALMPOERRRSDRRDSQGSVASTHWSDSSEVVN 451
 Db 641 VNISWVLPKNESGYDKFKGRTYWNLTSLNVESEGEFNLITDSRG-----SDSAGTLTQ 693
 QY 452 PYAEYGGARNSLSAHQPEEH--IYDEVAADPG---YSVIQ-----NFSGSGPVTGRL 498
 Db 694 PYNLNGISFNKDTTFNVARNARVNFIDIKAPIGINKYSSLNYSFNGNISVSGGSGVDFTL 753
 QY 499 IG-----TPCGGIQSTYALLANGSLGRLGMLTSGGETAVSSVNAAPTGPV 546
 Db 754 LASSNVQTPGVVINSKYFNVSTGSLRFTSGSTKFTGSIEKDLTLNATGGNI 807

RESULT 15
 US-08-302-832-2
 ; Sequence 2, Application US/08302832
 ; Patent No. 5603938
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; STREET: Bldg. 1
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/302.832
 ; FILING DATE: 16-SEP-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9205704.1
 ; FILING DATE: 16-MAR-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US pct/us93/02166
 ; FILING DATE: 16-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berkstresser, Jerry W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-404
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0813
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match 4.28; Score 117.5; DB 1; Length 1536;
Best Local Similarity 19.7%; Pred. No. 0.23;
Matches 129; Conservative 73; Mismatches 239; Indels 213; Gaps 26;

Qy 37 LISSTGALGSRSLFPLRNSMADSVDSRDIPGLPTNPRLAAATSETCLLGGF----- 89
Db 223 VISVNG--GSISLLAGQKIYISDIIN-----PTIYSIAAPENEAVNLGDIFAKGGNI 273
Qy 90 -----EVLHDKGLDILNTQIGPSAFRVEVQADGTHAAIGE----- 125
Db 274 NVRAATIRNOGKLSADSVSKDKNIVLSAKEGEAEIGGVISAQNOQAKGKLMITGDKV 333
Qy 126 -----KNGLEV-----SVTLSPQE-----W 140
Db 334 TLKTGAVIDLSGKEGGTYLGGDERGEGKNGIQLAKTISLEKSTINVSKEKGGRAIW 393
Qy 141 SSLQSID-----TEGKNRFVFTGG--RGGSGHPMTVVASDIAEARTILAKLDPDN----- 189
Db 394 GDIALIDGNINAQSGGDIAGTGGFVETSGHDLFKDNAIVDAKEWL---LDFDNVSINAE 450
Qy 190 HGGRPKQDVTRSVGVG--SAGSIDDGVVSETHSTTSSVRSDDPKFW----- 235
Db 451 TAGRSNTSEDEYTGSGNSASTPKRNEKKTTLTNTTLESILKKGTFVNTANQRIYVNSS 510
Qy 236 --VSVGATAAGLAATGAQALALTPEDDPTTDPDQAAANAESATKDQLT----- 287
Db 511 INLSNGSLTLWSEGRSGGV-----EINNDITGDDTRCAN-----LTIYSGW 554
Qy 288 QEAFKPNENKVNIDANGNAIPSGELXDDIVEQIAQAKAKEGEVAROQAVESNAQAQRY 347
Db 555 VDVHKN-----ISLGAQONI-----NITAKQDIAFEKGSNQVITGQG----- 591
Qy 348 EDQHARQEEQLSSGIGYGLSSALIYAGGIGAGVTTALHRRNQA-----EQTTT--- 398
Db 592 -----TITSGNQKQGFNNVSLNGTSGSLQFTTKRNTKYAITNKEGTLNISGK 640
Qy 399 TTTHTVVOOQTGGIPOHK-----VALMPOERRRFSRRDSQGSVASTHWSDSSESVN 451
Db 641 VNISMVLPKNEGYDKFKGRTYWNLTSLNVSESGEFNLTIDSRG-----SDSAGTLTQ 693
Qy 452 PYAEVGGARNLSAHQPEEH--IYDEVAADPG---YSVIO-----NFGSGPVTGRL 498
Db 694 PYNLNGISFNKDTTFNVERNARVNFIDKAPIGINKYSSLNYSFNGNISVSGGSDVFTL 753
Qy 499 IG-----TPQGGIGTSTYALLANSGLRLGMLTSGGETAVSSVNAAPTGPV 546
Db 754 LASSSNVQTPGVVINSKYFNVTGSSSLRFKTSKTKTGFSTIEKDLTLNATGNI 807
```


404 VVOQGGIPIHVKALMPQRRFRSDRRSQSGVASTHWSDS-SEVVNPYAEVGGARNS 462
Db VNK PANNTPAQGNVDTPGSDTMSRRSRSMASSTSTFTDTSIGVYQNPYADV-----K 462
463 LSAHQPEEHYDEVAADPGVSVIQNFSGSPVT-----GRLGTPOGGIO 507
Db TSLLD-----SQPFTSNWTSVQNMGNDSVYSTIQHPRTDITNGARLLGNPSAGIQ 516
464 STYALLANSGLRLGMGLTSGGETAVSSVNAAPTGPVRV 549
Db STYARLALSGGLRDMGLTGGTGSNSAVNTSNNPPAPGSHRFV 558
RESULT 2
A28182
hemolysin A - Serratia marcescens
C:Species: Serratia marcescens
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Oct-1999
C:Accession: A28182
R:Poole, K.; Schiebel, E.; Braun, V.
J. Bacteriol. 170, 3177-3188, 1988
A:Title: Molecular characterization of the hemolysin determinant of Serratia marcescens
A:Reference number: A28182; MUID:88257037
A:Accession: A28182
A:Molecule type: DNA
A:Residues: 1-1608 <POO>
A:Cross-references: GB:M22618; NID:g340726; PIDN:AAA50323.1; PID:g556420
Query Match 5.4%; Score 150.5; DB 2; Length 1608;
Best Local Similarity 20.9%; Pred. No. 0.25;
Matches 122; Conservative 84; Mismatches 236; Indels 141; Gaps 25;
405 LGGPEVLHDKGLDILNTQIGPSAFRVEQA-----DGFH----- 120
Db VNAAGDILDAQSDQSESR---SGFNKASAKGGFTADSKNFGAGFGGTHNGESSST 1122
406 AAIGEKNGLEVSVTLSPQEWSSLSQIDTEGKNNRFVFTGGRGSGHPVTVASDIAEA-RT 179
Db AQVGNISSG-QQGVELKAGRLTLQGTVDKSGDVSLSAGN-----KVALQAAESTQT 1173
407 RLAKLDPNHHGRQPKVDTRSVGVSAGS-IDDGVVSETHSTTNSVSRSDPKFWVS 238
Db RKESLSNIDLGAGSSDSKEKTNLSAGGAFDIKYNESATERQGIATSDGKVTLSA 1233
408 ---GATAAGLAGLAATGIAQALALTPEDDPTTDPDQAANA-----ESATKDQLTOEAFK 292
Db NGKGDALHLQAKVSGGSAAL-----BAKNGGILLESKANEQ----- 1271
409 NPEN-----QKVNIDANGNAP-----SGELXDDIVEIAQAK-----E 327
Db HKDNWSLGIKAKAGGQTFNKDAGGVDPNTGKDTHTLGAGLVKVGVEQDKTTHANTGIT 1331
410 AGEVARQQAQVESNAQAQRYEDOHARRQBELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387
Db AGDVTLSNKDTRL-AGARVD-----ADSVQGVGGDLHVESKRDVGVKVDVDAGLS 1384
411 RRNPQAEOTTTTTTHVVOQOCTGPIQ-----HKVALMPQR-----RRFSRRDSQGS 436
Db HSNDCSSITSLKSVGTGPYAGKVKLEAGVKNVADATTDKYNVARRLPDQDDTTGA 1444
412 VASTHWSDSSEVVNPYAEVGGARNSLAHQPEEHYDEVAADPGVSVIQNFSGSPVTG 496
Db VS-----FSKAEKGVTLTPATG-----EKPGPLWDRGARTVGVGAKVSDITGPAGRG 1493
413 RL-----LGTPOGGIQTSTYALLANSGLRLGMGLTSGGETAVSSV 531
Db HLKYNADVNNNAVGEQSA---IAGKNGVALQVGGQTQLTGE 1533

3
C71513
hypothetical protein CT456 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: C71513
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: C71513
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1005 <ARN>
A:Cross-references: GB:AE001319; GB:AE001273; NID:g3328881; PIDN:AAC68056.1; PID:g332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
C:Gene: CT456
Query Match 5.4%; Score 150; DB 2; Length 1005;
Best Local Similarity 22.8%; Pred. No. 0.14;
Matches 145; Conservative 65; Mismatches 234; Indels 192; Gaps 33;
414 GNLGNVNGNHLIPAPPPL-PSQTDGAARGGT-----GHLISSTGALGSRSLFSLR-- 54
Db GNLPTVTIINKEKTCVAYGPMNSQEAASGYTPSAMRRGHRVDFGGIFEKANDFNKINWG 490
415 ---NSMADSDV-SRDIPGLPTNPRLAAATSETCILLGFEVLHDKGLDILNTQIGPS 108
Db TQAGPSEDDGIFSNETPG--AGP---AAPSP-----PSSIPIN----- 528
416 AFRVEQADGTHAAIEKNGLEVSVT-LSPQEWSSLSQIDTEKNNRFVFTGGRGSGHPM 167
Db ---VNVVGTNVNIGDT---VNTTNTTPTQSTDASTDT----- 563
417 VVASDTAEARTRILAKLDPNHHGRQPKVDTR---SVGVG-----SASGIDDG 214
Db ---SIDDINT-----NNTDDINTDKSDGAGGVNGDISETESSGDDSG 607
418 VVSEHTSTTNSSVSRDPKFWVS-----GAIAAG-----LAG 247
Db SVSSSE-SDKNASVNGDGPAMKDILSAVRKHLQVVPYVPENGSGSTEGPLPANQTLGDVID 666
419 LAATGIAQALALTEP---PDDPTTTPDQAAANAESAT-----KDOLTOEA----- 290
Db VENKGSQDTKLSGNTGAGDDDPFTT---AAVNGAEETILSDTSGIGDDVSDTASSGD 724
420 ---FKNPENOKVNIIDANGNAIPSGELXDDIVEIAQAKAEAGEVARQQAQVESNAQAQR 346
Db EGGVSPSSSSSNKNTAVNDGPGSL---DILAARKHLDKVYPCDNGSGSTEGPLQANQT 781
421 QHARRQBELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEOTTTTT 401
Db LGDIVQDMETTGTQETVVSF--WKGSTSTESAGSGSVQT---LLFSPPTPTTTTLRT 837
422 ---VVOQOCTG-IPQHKV--ALMPQRRFRSDRRSQSGVASTHW 442
Db GTGATTTSLMMGGPIKADIITGGGRIPGGGTLEKLPRLRAHLDISFDAQGDVSTEE 897
423 SDSSEVVNPYAEVGGARNSLAHQPEEHYDEVAADPGVSVIQNFSGSPVTGRLIGTP 502
Db POLGS-IYNKFRQETGSRGILA-----FVESAPKPGSA--QVLTGTG-----GDK 940
424 GGGIQTSTYALLANSGLRLGMGLTSGGETAVSSV 538
Db GNLFOAAAATQALGNV-AGKVNLAIOGQKLSLVN 975
RESULT 4
D85644

C:Genetics:
A:Gene: CESP-T06D8.1
A:Map position: 2
A:Introns: 1391/3; 1432/3; 1470/3; 1505/1; 1520/1; 1616/1; 1644/1; 1687/3; 1742/1

Query Match 5.2%; Score 146; DB 2; Length 1829;
Best Local Similarity 21.7%; Pred. No. 0.58;
Matches 118; Conservative 63; Mismatches 256; Indels 108; Gaps 19;

QY 68 GLPTNPSRLAAATSETCLGGFEVLHDKGPLDILNTQIGPSAFRVEQADGTHAAIGEN 127
DB 172 GEETTTSAVTEASSEA-----TTTPAGTEASGEET-----TTSVTEGS 210
QY 128 GLEVS-----TLSPQWSSLSQIDTE-GKNRFVFTGGGSGHPMTVASDIAEARTRI 181
DB 211 GEETTVVAVVSSGEEEPASSSTSIPTLSKNDQV-----TEASGEETITAAATEASETTT 266
QY 182 LAKLDPNHHGROPKVDTRSVGVSAG-----IDGVVSETHTSTNNS 227
DB 267 SAVTEGSGE-----DTTWAVVVELSGEQPASSSTSIPTLSKDDQVTEASGEETTTA 318
QY 228 VRSDPKFWVSGAAGLAGLAATGIAQALALTPEDDPTTDP-----DOANAESA 281
DB 319 AATEASEETTTSVTEG-SGEETTVAVVSSGEEEPASSSTSIPTLSKDDQVTEASGEE 377
QY 282 TKDQLOTFAPKNQKVN-----IDANG-----NAIPSGELX--DDIVEQ 320
DB 378 TTTAAATEASEETTTSVTEGSGEDTTVAVVSSGEEEPASSSTSIPTLSKDDQVTEA 436
QY 321 IAQAKAGEVARQAVESNAQAQORVEDQ-----HARQEEQLSSGIGYGLSSALIV 374
DB 437 SGEETTTAAATEASEETTTSVTEGSGEDTTVAVVSSGEEEPASSSTSIPTLSKDDQV 496
QY 375 AGGICAGVTTALHRRNOPAEQTTTT-----TTHVVOOQTGGIPQHKVALMPOERR 425
DB 497 TEASGEETTTA-AATEASEETTTSVTEGSGEETTVAVVSSGEEEPASSSTSIPTLS 554
QY 426 RFSDRDQSGSVAST-HWSDSSSESVNPIAEVGGARNSLSAQHEEHIYDEVAADPGYSV 484
DB 555 KDDKYTEASGETTTTAAATDASSETTTTSVTEGSGEETTVAVVSSGEEEPASSSTSI 613
QY 485 IQNFS-----GSGPVTRLGIPGQGIQSIQIYALLANSGLRLGMGLTSGGETAVSSV 537
DB 614 PTELSKDDQVTEASGEETTTAAATEASEETTTSVTEGSGEETTVAVVSSGEEEPASS 673
QY 538 NAAPT 542
DB 674 TSIPT 678

RESULT 7
T13936
collar protein isoform C - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T13936
R:McGinnis, N.; Ragnhildstveit, E.; Veraksa, A.; McGinnis, W.
Development 125, 4553-4564, 1998
A:Title: A cap 'n' collar protein isoform contains a selective Hox repressor function.
A:Reference number: 217817; MUID:98453413
A:Accession: T13936
A:Status: preliminary;
A:Molecule type: mRNA
A:Residues: 1-1296 <MC>
A:Cross-references: EMBL:AF070064; NID:g3859888; PID:g3859889; PIDN:AAC72898.1
C:Genetics:
A:Gene: cnc
A:Cross-references: FlyBase:FBgn0000338
A:Map position: 3
C:Keywords: leucine zipper

Query Match 5.2%; Score 144.5; DB 2; Length 1296;
Best Local Similarity 19.9%; Pred. No. 0.44;
Matches 149; Conservative 68; Mismatches 213; Indels 319; Gaps 31;

QY 8 NNVNGNHLIPAPPLPSQ-TDGAARGGTGHLISLSTGALGSRSLFSLRNSMADSVDSRDI 66
DB 24 HNTTGNSSVQTAALQDVOSTSAAATGAT--MVVGTG----- 57
QY 67 PGLPNTPSRLAAATSETCLGGFEVLHDKGPLDILNTQIGPSAFRVEQADGTHAAIGEK 126
DB 58 -CAPTSSGQTSGA-----LG--EIHIDTASLDPGNANHSPLHTSELDFTFLHALQDQ 109
QY 127 NGI-----EVSVTLSPQWSSLSQIDTEGK-----RFVFTGGGSGHP 166
DB 110 RSIWEONLADLYNDLSLOTSPYANLPLKDGQPOPSNSSHLDLSLALLHGFTGGSGAP 169
QY 167 MVTVASDIAEARTILAKLDPNHHGROPKVDTRSVGVSAGSIDD-----GVV 216
DB 170 LSTAALNDSTPHRNLGVS--TNNAGR--SDGEEESLYLGRLFGEDEDEYEGELVGVGA 226
QY 217 S--ETHSTTNSVRS-----DPRFWVSVGAIAAGLAG 247
DB 227 NACEVEGLTTDEPFSGNCFANEVEIGDEEESIAEVLYKQVDYDLGFLDQEI--ING 283
QY 248 LAATGIAQALALTPEDDPT--TTDPD-----QAANAASATKD-----Q 285
DB 284 SYASGNSAATNVKSKPEDETKSSDSPISSESSGFKDQDVNAENASAAVDDIEKLKALEE 343
QY 286 LTQAFAPKNQKVNIDANGNATP-----SGE-----LXD 315
DB 344 LQDKKKNENQLEDITNEWNGIPFTIDNETGEYIRLPDELNDVLKSEFFLQDDLSN 403
QY 316 DIYEOTAQAKAGEVARQAV-----ESNAQAQORVE----- 348
DB 404 DPVASTSQAAAAFENQAQRIIVSETGEDLLSGEISSKQNRNEAKNDNDPEKADGDSFS 463
QY 349 -----DQHARRQEELQLSS----- 362
DB 464 VSDPEELQNSVGSPFLDDEDAKDELMLQSTVPSVHHPHHPHHPHSHHHASHM 523
QY 363 -----GIGYGLSSALIYAGGIGAGVTTALHRRNOPAEQTT 397
DB 524 HAHAAHAAAAAHQRAVQAQNGYGGVGVG-----VGVGVSGTGSFAQR--QPA----- 573
QY 398 TTTTHTVVOOQTGGIPQ-HKVALMPQERRRFSRRDSQGSVASTHWSDSSESVNPIAEV 456
DB 574 -----AGFHHCHGQRMPLNRSVSMERLQD---FATYFSPIS--MV 612
QY 457 GGARNSLSAQHEEHIYDEVAADPGYSVIQNFSGSPVTGRLIGTPGQGIQSIQIYALLANS 516
DB 613 GG-----VSDMSPPYPHY-----PGYSYQASPSNGAP-----GTPGQHGQY----- 648
QY 517 GGLRLGMGLTSGGETAVSSVNAAPTGP 545
DB 649 -----GSGANATLQPPPPPP 664

RESULT 8
S04722
puff 74E protein - fruit fly (Drosophila melanogaster)
N:Alternate names: ecdysone-induced protein E74B; ets-related protein E74B
C:Species: Drosophila melanogaster
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000
C:Accession: S04722; B34692
R:Janknecht, R.; Taube, W.; Luedicke, H.J.; Pongs, O.
Nucleic Acids Res. 17, 4455-4464, 1989
A:Title: Characterization of a putative transcription factor gene expressed in the 20
A:Reference number: S04722; MUID:89315191
A:Accession: S04722
A:Molecule type: DNA
A:Residues: 1-883 <JAN>
A:Cross-references: EMBL:X15087; NID:g7513; PIDN:CAA33195.1; PID:g7514
A>Note: Gln-867 was also found

R:Burtis, K.C.; Thummel, C.S.; Jones, C.W.; Karim, F.D.; Hogness, D.S.
Cell 61, 85-99, 1990
A:Title: The *Drosophila* 74EF early puff contains E74, a complex ecdysone-inducible gene
A:Reference number: A90912; MUID:9019900
A:Accession: E34692
A:Molecule type: DNA
A:Residues: 1-866,'Q',868-883 <BUR>
A:Cross-references: GB:M37083; NID:q157309; PID:q157310

A: Gene: FlyBase:Elip74EF
 A: Cross-references: FlyBase:FBgn0000567
 A: Map position: 3 74EF
 C: Superfamily: ets DNA-binding domain homology
 C: Keywords: alternative splicing; DNA binding; transcription regulation
 F: 789-869/domain: ets DNA-binding domain homology <ETS>

| | | | | |
|-----------------------|--------------|-----------------|-----------------|-------------|
| Query Match | 5.1% | Score 144; | DB 2; | Length 883; |
| Best Local Similarity | 20.8% | Pred. No. 0.28; | | |
| Matches 124; | Conservative | 68; | Mismatches 197; | Indels 208; |
| Gaps 26; | | | | |

| | | | |
|----|-----|---|--------|
| Qy | 39 | SSTGALGRSLFSPURNMADSVDSRITPGLPNPSRLAAATSETCLGGFEVLHD---- | 94 |
| | | : : : : : | |
| Db | 38 | SSSUSLSSSSSSLSUUAATPTPVAS---PVTPTSPPAAAAPAEASPAGAEQLQDGOQA | 94 |
| | | : : : : : | |
| Qy | 95 | ---RGPL-----DI LNTQTIGPSAFREVQADG----THAAIGEKNGLSEVLTSPQEWSSL | 143 |
| | | : : : : : : : : : : : : : : : : : : | |
| Db | 95 | KTQEDPTMKQODMLEKTRQEYKDPVNVEEPGAIVDTESVMARQSPPVASTKYPE--- | SL 151 |
| Qy | 144 | QSIDTEGNKRVFTGGRRGSGHPMV-----TVASIDIAEARTILAKLDPDNDHGGQP | 195 |
| | | : : : : : : : : : : : : : : : : : : | |
| Db | 152 | EELSNKS-----PPVOEDEERSESVASDCREFKVLYNLHLCROOHH--HSP | 194 |

QY 196 KDVD-TRSVGSGSAGIDD--GVVSETHSTNTSSVRSDPKFVWSVGAIAAGLAGLAATG 252

QY 253 IQAALATPE-----PDDPTTDPDQANAAESATKDQLTQFAKPNQKVNIDA 303

Db 224 ITAAPTLPQQHQOPMSDIEDEETLEDVDDADADVEADAEDEELLEQQYONGYDPSPLDLSL 283

Db 284 GGATSAAGAAAAAASAVRRRRGRITYSGTESDDSAQCERARMLKPKPERKAERSAAYKKSLM 343

Qy 339 -----SNAQAQRVEDQHARRQEELQLSSGIGYGLSSALIVAGGIGAVTT 384

344 KVVVETETDVTUKCSFSDAPDQQLQQQQHQQQQQQQQQQQQQQQQQQQQQQQQ 390

nh -----ACACAC 390

Qy 385 ALHRRNPQAEQTTHHTVVQOQTGGIPQHKVLMQPERRRFRDRRDSQGSVASTHWS 444

| | | | | | |
|----|-----|------------|--------------------------|-------------|-----|
| DB | 391 | LLHIK--- | TEQNLLTPQLQQQQ | -----QQQQG | 416 |
| Ov | 445 | SSSEVVNPYA | EVGGARNLSAHOPEEHYDEVAAPG | SVIONFSGGPV | 494 |

Db 417 ----LHGAAGNGSSNGNAHQOQPL-----AIPQRPLLHLLSGGATHNPHRNYTT 466

Db 467 ATTTGSPPPPADSGVSDVS--SSGGGPCCADELKGRLGMPATASAAAAA 521

RESULT 9
T42233 submaxillary mucin 1 - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Nov-2000
C:Accession: T42233
R:Jiang, W.; Woitach, J.T.; Keil, R.L.; Bhavanandan, V.P.
Biochem. J. 331, 193-199, 1998
A:Title: Bovine submaxillary mucin contains multiple domains and tandemly repeated, non
A:Reference number: Z22092; MUID:98180948
A:Accession: T42233

A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1589 <JIA>
A;Cross-references: EMBL:AF016599; NID:g3057086; PID:g3057087; PIDN:AAC3935
C;Genetics:
A;Gene: SM1
C;Superfamily: pig submaxillary mucin

Query Match 5.1%; Score 144; DB 2; Length 1589;
Best Local Similarity 19.5%; Pred. No. 0.63;
Matches 119; Conservative 87; Mismatches 247; Indels 156; Gaps

| | | | | | |
|----|-----|---|---------------------------------|-----------------|-----|
| Qy | 7 | GNVNVGNHLLPPAPPPLPSQTDGAA | -----RCGTGCHLISS----- | TGALGSRSLFSPURN | 55 |
| | | | | | |
| Db | 66 | GTNVP-----VSGAPVTPGSSAGSGAPGTGGPGSETASPLSGAAGTSATGKRTSTIPP | ----- | | 118 |
| | | | | | |
| Qy | 56 | SMADSVDSRDIPLGPNPRLAAATSETCL | -----LGGFEVLHDKGPLILNTQTGPS---- | | 108 |
| | | | | | |
| Db | 119 | --GGAPVTPPPLLIITGASAGAPPASSSEVTIPLPGATGTDVLRSGTSLPVSQGVATVPASSP | 176 | | |

| | | | | |
|----|-----|---|--|-----|
| Qy | 109 | ---- | --APRVDVQA-----DGHAAIGENKNGLEVSVTLSPQDKWSSLQ | 144 |
| Dd | 177 | GGSSATAGPAVGOTTVVQVSGAFATHVKKASNTGSSAEISSETTGATAGTLLTSE--TSIV | 234 | |
| Qy | 145 | SIDTEKNREV-----FTGGRGGSQHPPMVTVASDIAEARTILAKLPDN-HG | 191 | |
| Dd | 235 | SSATRAPSAAVTRPVTYDDVGISHSSSRRGRRTIVIGPSSVSVAEQIAPSLSTPDGLEG | 294 | |
| Qy | 192 | GROPKDVDTRSV----GVGSAAGIDCGVYSETHTTNNSVSRSDPKFVWVCIAAAGLAG | 247 | |
| Dd | 295 | TTKISLDVDAARTIRPSYGALCATGSSIGEIGTSTS-----PEF-TETWSFSFVGUR- | 343 | |

QY 248 LAATGIAQALITPEDD-PTTTDPDQAANAESAATKDQLTQEAFFKPNQKVNIDANGN 305

QY 307 AIP-SGELXDDIVEQTAAQAKEAG-----EVARQAVESNAQAQQRYEDQHARR 354

Db 392 TRPSSGETETTVIESRVSGSDEGLGTIGTAGLMRTTRISVVVSGTGPSSGETGSA -- 449

Db 450 VSEFRTSGSLGKG-SETTVSTPGL-ARMTRISFGGSRTRTQSSGETGTTVIESRTGSP- 506

[illegible]

QY 475 EVAADPGYSVIONFSGSGPVTGRLLIGTPGGTQSTYALLANSGLRLGMGG ---LTSGG 530

DB 547 -----SRNGSLGEGSGTTGAIAGLTRTRTRLSWVGSGTTRPSSG 585

Ov 531 ETAVSSVNA 539

Db 586 ETRTIVS 594

RESULT 10
T39141
hypothetical protein SPAC8C9.04 - fission yeast (*Schizosaccharomyces pombe*)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39141
R:Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z21748
A:Accession: T39141
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-647 <OLI>
A:Cross-references: EMBL:Z99168; PIDN:CAB16292.1; GSPDB:GN00066; SPDB:SPAC8C9
A:Experimental source: strain 972h-; cosmid c8c9

C:Genetics:
A:Gene: SPDB:SPAC89.04
A:Map position: 1

Query Match 5.1%; Score 141.5; DB 2; Length 647;
Best Local Similarity 19.0%; Pred. No. 0.25;
Matches 103; Conservative 64; Mismatches 207; Indels 169; Gaps 20;

QY 114 VQADGTHAAIGKNGLEVSTLSPQWSSLSQSDTEKKNREFTVGTGGGSGHPMTVAD 173
DB 94 LQEPGFTQIVKDAQDQVDEPLEPIASSALGTVEPTDNK-----PSASTSTA 141
QY 174 IABARTILAKLPDNGHQKQKDVTRSVGSGASIDGQVSETHSTTN----- 225
DB 142 VPTTEARTSITEPANSPSSSSSSASTKATQSA---DYVVAEHFAPORNDQLGNSP 198
QY 226 SSVSRDPKFWVSV---GAIAAGLAGLAATGIAQALALTPP---PDDP-----TTTDPD 272
DB 199 ASITSKPAATTSAAQPSKVENAKATSQPIITAEKEIPELKPIEPAIMISKEINTHD 258
QY 273 QAA-----NAAESATKQLTQEAQFN-----P 294
DB 259 QAAATTTAAVASASTTATAESHAVADGIMDNVLESIGENVQOETVFEEDASDIPHADVIP 318
QY 295 ENQKVNID-----ANGNAIPSG---ELXDDIVEQIAQAQAKEAGEV 331
DB 319 HTTTVTVEEPIALGOGVYTHEATTSARASASGIPGAFEEVQOTVOEDLPHPTAEIVEI 378
QY 332 ARQAVESNAQAQORYEDQHARRQEELQLSGSIGYGLSSAL-----IV 374
DB 379 ARFAEQPVRAQQPEYESSVV---QEAETVTDVCKGVSSYTKNEVNPSTIPTSESNPVA 436
QY 375 AGGIGAGVYTTALHRRNQ-----PAE-----QTTTTHTVVQOQTGGIPQHKVAMQER 424
DB 437 VGG-----TTAEHPVQEAHTAPTETAHDFSKETTATSKRVSKHDKASAEKHVKARKP--- 488
QY 425 RFSDRRDSQGSVASTHWSDSSESVNVNPEYEVGGARNLSAQHPHEEHIYDEVAADPGYSV 484
DB 489 -----SSTGQEPPTSTPAKSNOSKHHARRSKQ-----ASAP----- 521
QY 485 IQNFGSGPVTGRLLIGTPGQGIQSTYALLANSGLGLMGGLTSGGETAVASVNA----- 541
DB 522 ----SSPGTTSAAVPGKKSAIEAA-ADIPTSADTVESKHAAGSGSATTPSPGSATTKP 576
QY 542 TPG 544
DB 577 TPG 579

RESULT 11
A55182
aggrecan precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jan-2000
C:Accession: A55182; S55329; S50207; S51355; I78532; I58123
R:Walcz, E.; Deak, F.; Erhardt, P.; Coulter, S.N.; Fuelleop, C.; Horvath, P.; Doege, K.J.
Genomics 22, 364-371, 1994
A:Title: Complete coding sequence, deduced primary structure, chromosomal localization,
A:Reference number: A55182; MUID:95104847
A:Accession: A55182
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-2132 <WAL>
R:Watanabe, H.; Gao, L.; Sugiyama, S.; Doege, K.; Kimata, K.; Yamada, Y.
Biochem. J. 308, 433-440, 1995
A:Title: Mouse aggrecan, a large cartilage proteoglycan: protein sequence, gene structure
A:Reference number: S55329; MUID:95289972
A:Accession: S55329
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-567, 'G', 569-1728, 'I', 1730-2132 <WAT1>

A:Cross-references: GB:U22901; NID:g886014
R:Glumoff, V.; Savontaus, M.; Vehanen, J.; Vuorio, E.
Biochim. Biophys. Acta 1219, 613-622, 1994
A:Title: Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues
A:Reference number: S50206; MUID:95035091
A:Accession: S50207
A:Molecule type: mRNA
A:Residues: 350-481, 'R', 483-506 <GLU1>
A:Cross-references: EMBL:X80279; NID:g673432
R:Glumoff, V.
submitted to the EMBL Data Library, July 1994
A:Reference number: S51355
A:Accession: S51355
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 350-383, 'CPVMSQSRPMAA' <GLU2>
A:Cross-references: EMBL:X80279
R:Watanabe, H.; Kimata, K.; Line, S.; Strong, D.; Gao, L.Y.; Kozak, C.A.; Yamada, Y.
Nature Genet. 7, 154-157, 1994
A:Title: Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the agg
A:Reference number: I58123; MUID:95004579
A:Accession: I78532
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 211-326 <WAT2>
A:Cross-references: GB:S73722; NID:g765215; PIDN:AAB32160.1; PID:g765216
A:Accession: I58123
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 211-240, 'MCTASLRWRVRSFMRHPQNRSPSRQPT', 'AGGWHAWPQASSTWPGRAVWTCALAGW'
A:Cross-references: GB:S73720; NID:g765211; PIDN:AAB32159.1; PID:g765212
C:Genetics:
A:Map position: 7
A:Introns: 253/1
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
C:Keywords: cartilage; extracellular matrix
F:1-19/Domain: signal sequence #status predicted <SIG>
F:44-135/Domain: immunoglobulin homology <IMM>
F:170-247/Domain: link protein repeat homology <LNK1>
F:268-349/Domain: link protein repeat homology <LNK2>
F:504-581/Domain: link protein repeat homology <LNK3>
F:602-683/Domain: link protein repeat homology <LNK4>
F:1922-2042/Domain: C-type lectin homology <LCH>
F:2049-2105/Domain: complement factor H repeat homology <FHD>

Query Match 5.0%; Score 140; DB 1; Length 2132;
Best Local Similarity 20.5%; Pred. No. 1.7;
Matches 129; Conservative 81; Mismatches 241; Indels 178; Gaps 29;

QY 4 GNIGNNVNHLLIPAPPLPSQTDGAARGGTGHLTSTGALGSRSLFSLRNSMADSVDS 63
DB 906 GDLGLPSGGITETSTSGAETSGLPSGGD-LETSTSGVDDVSGIPTGRELETSASG 964
QY 64 -RDIPGLTNPRLAAATSETCLGGFVLHDKGLDILNTQIGPSAFRVEVQADGTHAA 122
DB 965 VEDLSGLPSGEGSETSTSGIE-----DISLPTG-GES-----LETSSAG----- 1004
QY 123 ICEKNGLEVSTLSPQWSSLSQSDTEKKNREFTVGTGGGSGHPMTVAD-----DIAEART 179
DB 1005 VGDLGL-----PSGESLETSSAGAEVDVQLPTERGG-----LETSSAGVEDITVLP 1053
QY 180 RILAKLDPNHHGGRPKDVTDSVGVSGASGIDGQVSETHSTTSSVSRDPKFWVSVG 239
DB 1054 -----GRE--SLETSSAGVEDVSGLPSPG---REGLETSSAGIEDISVFPTEAE 1096
QY 240 AIAAGLAGLAATGIAQALALTPEDDPTTDP-----DOAANAES----- 280
DB 1097 GLDTSASGGVSGI-----PSGGDGTSTASGVEDVSGLPSSGEGLETSSAGVEDLGP 1149
QY 281 AYKDKLTQEAQFNKPNQKYNIDANGNAIPSG-----LXDDIVEQIAQAQAKEAGEV 331
DB 1150 STRDSL-----ETSSAGVDVTG---FFSGRCDPSTSVSGVGGDF-----SGLP 1189

[illegible]

QY 483 SVIQFSGGPVTGR----LIGTPGGIQSTYALLANSGLRLGMG 524
|: || | : | : || | : : | || | : : |
Db 1189 SM-----GSMPTSFPRGALVGTAAQEKRTILVENAPSGYLKISSG 1229

Search completed: September 27, 2001, 14:21:42
Job time: 78 sec

THIS PAGE BLANK (USPTO)

| Result No. | Score | Query | | Length | DB | ID | Description |
|------------|-------|-------|------|--------|------------|--------|--------------|
| | | Match | | | | | |
| 1 | 157.5 | 5.6 | 494 | 1 | SX9_CHICK | P48434 | gallus gall |
| 2 | 150.5 | 5.4 | 1608 | 1 | HLVA_SERMA | P15320 | serratia ma |
| 3 | 144.5 | 5.1 | 883 | 1 | E7AB_DROME | P11536 | drosophila |
| 4 | 140 | 5.0 | 2132 | 1 | PGCA_MOUSE | Q61282 | mus musculus |
| 5 | 137 | 4.9 | 1532 | 1 | IGA_NEIGO | P09790 | neisseria g |
| 6 | 136 | 4.8 | 1569 | 1 | YPYA_SCOTI | P52143 | escherichia |
| 7 | 135.5 | 4.8 | 2124 | 1 | PGCA_RAT | P07897 | rattus norv |
| 8 | 134.5 | 4.8 | 1150 | 1 | APMU_PIG | P12021 | sus scrofa |
| 9 | 132.5 | 4.7 | 1025 | 1 | SLAP_CAUCR | P35828 | caulobacter |
| 10 | 132 | 4.7 | 827 | 1 | CSG_HALYO | P25062 | halobacteri |
| 11 | 131.5 | 4.7 | 618 | 1 | ZEST_DROVI | Q24762 | drosophila |
| 12 | 130.5 | 4.7 | 940 | 1 | MAZ4_SCHCO | P37938 | schizophyll |
| 13 | 130 | 4.6 | 1260 | 1 | AL51_CANAL | P46590 | candida alb |
| 14 | 130 | 4.6 | 1596 | 1 | MAM_DROME | P21519 | drosophila |
| 15 | 130 | 4.6 | 2703 | 1 | NOTC_DROME | P07207 | drosophila |
| 16 | 129 | 4.6 | 2415 | 1 | PGCA_HUMAN | P16112 | homo sapien |
| 17 | 128 | 4.6 | 827 | 1 | XANP_XANS2 | Q60106 | xanthomonas |
| 18 | 127.5 | 4.6 | 509 | 1 | SX9_PIG | O18896 | sus scrofa |
| 19 | 127 | 4.5 | 784 | 1 | SP4_HUMAN | Q02446 | homo sapien |
| 20 | 126.5 | 4.5 | 2493 | 1 | Y5AA_USTWA | P49606 | ustilago ma |
| 21 | 125.5 | 4.5 | 673 | 1 | CY52_HUMAN | O60299 | homo sapien |
| 22 | 125 | 4.5 | 2442 | 1 | CBP_HUMAN | O92793 | homo sapien |
| 23 | 124 | 4.4 | 1589 | 1 | PHP_DROME | P33769 | drosophila |
| 24 | 123.5 | 4.4 | 829 | 1 | E74A_DROME | P20105 | drosophila |
| 25 | 123.5 | 4.4 | 2464 | 1 | MAPB_MOUSE | P14273 | mus musculu |
| 26 | 123.5 | 4.4 | 3591 | 1 | FTAB_BORPE | P12255 | bordetella |
| 27 | 123 | 4.4 | 980 | 1 | SYN_DROME | Q24546 | drosophila |
| 28 | 122.5 | 4.4 | 1306 | 1 | MSB2_YEAST | P32334 | saccharomyc |
| 29 | 122.5 | 4.4 | 2541 | 1 | TALI_MOUSE | P26039 | mus musculu |
| 30 | 120.5 | 4.3 | 314 | 1 | YR83_MYCTU | Q10691 | mycobacteri |
| 31 | 120 | 4.3 | 706 | 1 | HTF4_MOUSE | Q61286 | mus musculu |
| 32 | 120 | 4.3 | 995 | 1 | Y109_YEAST | P40442 | saccharomyc |
| 33 | 120 | 4.3 | 1505 | 1 | CDP_HUMAN | P39880 | homo sapien |

Db 292 DVNEFDQYLPNGHPGVATHQVVTYSGTYGISSASSPAGAGHAMMAKQPPQPPA 351
QY 269 TDPOAANAASATKDQLTQEAFAKNPENQKVNIDANGNAIPSGELXDDIVEQIAQAQKEA 328
Db 352 OPPAQ-----HTLPS-----TEREQGPQQR 372
QY 329 GEVARQAVESNAQAQRYEDQHARROEELQSSGIG-----YGLSSALIVAGGIGAV 382
Db 373 PHIKTEQLSPSHNSEQOQHPEQOQQOQQOQ-----GYGSFNLQHYGFSYPPPI-----419
QY 383 TTALHRRNQPAEQTTTTTHVVQOQGGIPQHKVQALMPQERRRFSRRDSQGSVAS--T 440
Db 420 -----TRSEYDYTEHQNSG-----SYISHAAGOSGSLYSTFT 451
QY 441 HWSDSSEVVNPNYAEVGGARNLSAHOPEEH 471
Db 452 YMNPTQRPMTPIADTSGVPTIPQTHSPQQH 482

RESULT 2

HLVA_SERMA
ID HLVA_SERMA STANDARD; PRT; 1608 AA.
AC P15320; 1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE HEMOLYSIN PRECURSOR.
CN SHLA
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RC STRAIN=SN8;
RX MEDLINE=86257037; PubMed=3290200;
RA Poole K., Schiebel E., Braun V.;
RT "Molecular characterization of the hemolysin determinant of Serratia marcescens."
RL J. Bacteriol. 170:3177-3188(1988).
CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY DEFINED.
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA REQUIRES SHLB FUNCTION.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M22618; AAA50323.1; -
DR PIR; A28182; A28182.
KW Hemolysis; Toxin; Outer membrane; Signal.
FT SIGNAL 1 30
FT CHAIN 31 1608 HEMOLYSIN.
SQ SEQUENCE 1608 AA; 165078 MW; D669B476FE7DAD51 CRC64;

Query Match 5.4%; Score 150.5; DB 1; Length 1608;
Best Local Similarity 20.9%; Pred. No. 0.42; 236; Indels 141; Gaps 25;
Matches 122; Conservative 84; Mismatches

QY 33 GPGHLISSTGALGSRSLRSPLRNSMADSVDSRDIPGLPTNPSPR-----IAAATSETC 84
Db 1008 GSDLIVDAKGEGGTORSNSASQAVTGSIDRAN--GINVNVKKDAIYQGTALNGRGRKTA 1065

QY 85 LLGGEVFLHDKGLDILNTOIGPSAFRVEVQA-----DQTH-----120
Db 1066 VNAGGDRLDQASDKQSESR---SGFNVKASAGGFTADSKNFCAGFGGCTHNGESSST 1122
QY 121 AATGKNGLEVTSLSPQWSSLSQSDTEGKNRFVFTGGRGSGHPMTVVASDIAEA-RT 179
Db 1123 AQVGNISG--QOQVELKAGRLDTLQGTVDKSGQDVSLAGN-----KVALQAESTQT 1173
QY 180 RILAKLDPDNHHGRQPKVDVTRSVGVGSASG-IDDGVVSETHSTTNSSVSRDPKFWVS 238
Db 1174 RKESKLSGNDLNGSSDSKEKTGGLNLSAGAFDIAKVNESATEROGATIASDGKVTLSA 1233
QY 239 ---GATAAGLAGIAAATGIAAALALTPEDDPTTDDPOAANAA---ESATKDQLTQEAFA 292
Db 1234 NGKGDALHLQGAQKVGSGSAAL-----EAKNGGILLSEAKNEQ-----1271
QY 293 NPEN-----OKVNIDANGNAIP-----SGELXDDIVEQIAQAQAK-----E 327
Db 1272 HKDNWSLGIKANKAGGQTFNKGAGKVDPTNKTHTLGLAGLVGEQDKYTHANTGIT 1331
QY 328 AGEVARQAVESNAQAQRYEDQHARROEELQSSGIGYGLSSALIVAGGIGACVTTALH 387
Db 1332 AGDVTLSNGKDTL-AGARVD-----ADSVQGVKVGDLHVESRKDVENGKVDVDAGLS 1384
QY 388 RRNPQAEQTTTTTHVTVVQOQTGGIPQ-----HKVALMPQER-----RRESDRDSQGS 436
Db 1385 HSNPDGSSITSKLSKVGTPRYAKVKEKLEAGVKNKADATTDKYNVNSVARRLDPOQDTTGA 1444
QY 437 VASPHWSDSSSESVNPNYAEVGGARNLSAHOPEEHIDEVAADPGYSVQNFSGSGPVTG 496
Db 1445 VS-----FSKAEKGVTLPATPAG-----EKPGQLMDRGARTVGGAVKDSITGPAGROG 1493
QY 497 RL-----IGTPCGQIGSTYALLANSGLRLGMGGLT--SGGE 531
Db 1494 HLKVNADVWNNNAVGEQSA-----IAGKNGVALQVGGQTQLTGGE 1533

RESULT 3

E74B_DROME
ID E74B_DROME STANDARD; PRT; 883 AA.
AC P11536; Q9VVI7;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ECDYSONE-INDUCED PROTEIN 74EFB (ETS-RELATED PROTEIN E74B).
GN EIP74EF OR E74B OR CG6285.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89315191; PubMed=2501755;
RA Janknecht R., Taube W., Lueddecke H.-J., Pongs O.;
RA "Characterization of a putative transcription factor gene expressed in the 20-OH-ecdysone inducible puff 74EF in Drosophila melanogaster."
RT Nucleic Acids Res. 17:4455-4464(1989).
RL [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=90199900; PubMed=2107982;
RX Burtis K.C., Thummel C.S., Jones C.W., Karim F.D., Hogness D.S.;
RA "The Drosophila 74EF early puff contains E74, a complex ecdysone-inducible gene that encodes two ets-related proteins."
RL Cell 61:85-99(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,


```
QY 332 ARQQAVERNAQAQRYEDQHARRQBELQLSSG----- 363
DQ 1190 SGKEGLETSASGAEADLSGLPSCKEDLVGSASGALFGLPPGTLGSGQTPEVNGPSPGFS 1249
QY 364 -----ICYGUSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHVVVQQTGGIP-QH 415
DQ 1250 GEYSAGIDGSGSPGLPDPFSGLPFPPTVSLVDSTLVEITATTSLEGRGTIGTSGSG 1309
QY 416 KVALMPOERRRSDRRDSOGSVASVHSDSSSEVVNPNYAEVGVGARNLSAHPHEHYDE 475
DQ 1310 EVSGLP--LGLDSSADISGLPSGTSLSQASGSPDSSGTS-----PFD 1353
QY 476 VAADP--GYSVIFNQFSGSPVTRGLRIGTGQGIQSTYALL-----ANSGLRLG 522
DQ 1354 VSGQPPGSGVSEETSGIPEISSQPSGTPDTTATSGVTNELNLSGQPDVSGDGSGLPFG 1413
QY 523 MG---GLTS-GGETA-VSSVNAAPTPGPV 546
DQ 1414 SQSSGITSVSGTSGISDLSQSPGFPV 1442

RESULT 5
IGA_NEIGO
ID IGA_NEIGO STANDARD; PRT; 1532 AA.
AC P09790;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE IGA-SPECIFIC SERINE ENDOPEPTIDASE PRECURSOR (EC 3.4.21.72) (IGA
DE PROTEASE).
GN IGA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=MS11;
RA MEDLINE=871115823; PubMed=3027577;
RA Pohlner J., Halter R., Beyreuther K., Meyer T.F.;
RT "Gene structure and extracellular secretion of Neisseria gonorrhoeae
RT Iga protease."
RL Nature 325:458-462(1987).
RN [2]
RP ACTIVE SITE.
RX MEDLINE=90154052; PubMed=2105953;
RA Bachovchin W.W., Plaut A.G., Flentke G.R., Lynch M., Kettner C.A.;
RT "Inhibition of Iga1 proteinases from Neisseria gonorrhoeae and
RT Hemophilus influenzae by peptide prolyl boronic acids."
RL J. Biol. Chem. 265:3738-3743(1990).
CC -!- FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.
CC -!- CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
CC CERTAIN PRO-I-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
CC SUBSTRATES ARE KNOWN.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE.
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04835; CAA28538.1; -.
DR PIR: A26039; A26039.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710; -.
DR PRINTS: PR00921; IGASERPTASE.
```

```
KW Hydrolase; Serine protease; Zymogen; Autocatalytic cleavage;
KW Transmembrane; Signal.
FT SIGNAL 1 27.
FT CHAIN 28 986 IGA-SPECIFIC SERINE ENDOPEPTIDASE.
FT PROPEP 987 1532 HELPER PEPTIDE.
FT ACT_SITE 278 278 POTENTIAL.
FT SITE 986 987 CLEAVAGE (AUTO-).
FT SITE 1018 1019 CLEAVAGE (AUTO-).
FT SITE 1121 1122 CLEAVAGE (AUTO-).
FT SITE 1532 1532 CLEAVAGE (AUTO-).
SQ SEQUENCE 1532 AA; 168976 MW; 68FF4112BD22F40D CRC64;

Query Match 4.9%; Score 137; DB 1; Length 1532;
Best Local Similarity 21.0%; Pred. No. 2,3;
Matches 102; Conservative 60; Mismatches 183; Indels 140; Gaps 20;

QY 34 TGHLSSTGALGSRSLFSPLRNSMADSVSDRIPGLTPNPSRLAAATSETCLIGGVEVUH 93
DQ 782 TGYVTCTGNLSDKAL-----NSFDATRINGVNLNQNAAVLGKAALWGKIQ--- 829
QY 94 DKGPLDLNLTQIGPSAFRYEVQADGTHAAIGKNGLEVSVTLSPQEWSSLSQSIDTSGKNR 153
DQ 830 -----GQGNRSRVSLNQHSHLTDGSDVHNSLADSHIHLNNAASDAQSANKYH 877
QY 154 FVFTGGRGSG--HPMVTVASDIAEARTILAKLPDNDHGGROPKDVDRSVGVGSASGI 211
DQ 878 TIKINHLNLSNGHFHYLTDLAKNLGD---KVLVK---ESASGHYQLHVQNT-GEPNQOGL 930
QY 212 D-DGVVSEHTSTTNSSVSRDPKFWVSVGAIAAGLAGLAAT-----GIAQALALPPEP 264
DQ 931 DLFDA-----SSVDRSRLFVSLANHYVDLALRYTIKTENGITRLYNPYAGNG 979
QY 265 DPTTTPDQAAANAASATKDLQTOEA-FKNPENQKNVIDANGNAIPSGELXDDIVQIAQ 323
DQ 980 RPKPAPSPAANTASQAQKATQTDGAQIAKNIVV-----APPSPOANQAEALRQ 1031
QY 324 QAKEAGEVARQOAVESNAQAQRYED-----QHARRQBELQLSSGIGYGLSALIVAGG 377
DQ 1032 QAK-AEQVKRQQAEEKVARQDEAKRKAEEIARQOEAR-----KAEELAANK 1081
QY 378 IGAVVTTALHRRNQPAEQTTTTTHVVVQQTGGIPQHKVALMPOERRR----- 426
DQ 1082 KAEERKARELARQKAEAS-----HQANAKPKRRRRRAILPRPPAPV 1124
QY 427 FS-----DRDSOGSVASTHSDSS-----SEVNPYAEVGVGARNLSAHPQEP 470
DQ 1125 FSLDDYDADKN-----SESSIGNLARVTPRMGRELINDYEEI-----PLE 1164
QY 471 HIYDE 475
DQ 1165 ELEDE 1169

RESULT 6
YFJA_ECOLI
ID YFJA_ECOLI STANDARD; PRT; 1569 AA.
AC P52143; P76610; P77017; P77019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 98.4 KDA PROTEIN IN ALPA-GABD INTERGENIC REGION (F949).
GN YFJA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
```

```

RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K.,
RA Kasai H., Kashimoto K., Kim S., Kimura S., Kitagawa M.,
RA Kitakawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H.,
RA Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Tagami H., Takemoto K., Wada C.,
RA Yamamoto Y., Yano M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).
CC -1- SIMILARITY: STRONG, TO BODETELLA PERTACTIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U36840; AAA79815.1; ALT_SEQ.
DR EMBL; AE000350; AAC75695.1;
DR EMBL; D90890; BAA16518.1; ALT_INIT.
DR EMBL; D90890; BAA16518.1; ALT_INIT.
DR EcoGene; EG13213; ypfA.
KW Hypothetical protein; Outer membrane.
SQ SEQUENCE 1569 AA; 162774 MW; 0D006A9A32154596 CRC64;
-----
Query Match 4.98; Score 136; DB 1; Length 1569;
Best Local Similarity 20.98; Pred. No. 2.7;
Matches 127; Conservative 69; Mismatches 259; Indels 154; Gaps 26;
QY 19 APPLSFQDTGAAGT--GHLISFGALGSLRSLFSLPLNSMADSVDSRDIPGLPTNPSRL 76
DB 579 AKHVQQGGGALIASTTSGTLIEGNSVGDAFY---IRNSEAKNV-----VLENAGSL 628
QY 77 AAATSETCLGGFEVLHDKPGLDLINTQIGP---SAFRVEQVADGTHAAIEKNGLEVS 133
DB 629 TVVTGSRV---DTIINANGKMDVYGDVGTVLNSAGTQTIYASATSKANIKGGKQTVY 685
QY 134 TLSPQE--WSSLQSIDTEKNEFVTGGRGSGHPWTVASDIAEARTRILAKLPD--- 188
DB 686 GLATEANIESGOIVDGGSTKTHNGGTQVQYNGKAIINTDIVSLQIQIMANGTAEGSI 745
QY 189 -----NHGGRQPKDV-----DTRSVGVGSASGI---DDGVVSETHSTTNSVRS 230
DB 746 INGGQVNEGGGLAENSVLNDGGTLDVRE--KGSATGIQSSQGAIVATRAVTVGTGRA 803
QY 231 DPKFWVSGAIALAGLAATGI--AQALALPEPDDPTTTPDQQAANAESATKDOLTQ 288
DB 804 D-----GVAFSEIQAANNILLANGVLAVESD--TSSDKTVQNMGGREIVTKAT- 852
QY 289 EAFKFNQKVNIDANGNAIPSCGLXDDIVEQIA-----QQAKEAGEVARQQ--- 335
DB 853 -----ATGTLITGGE--QIVEGVANETTINDGGIQIVSANGEAIKTKINE 895
QY 336 -----AVESNAQAQRYEDQHARRQEE-----LQLSSGIGYGLSS-----A 371
DB 896 GGTITVNDNGKATDIVQNSGAALQTSANGIEISGTHOYGTFSISGNLATNMLLENGNL 955
QY 372 LIVAGGIGAGVTTALHRRNPQAGOTTTTHTTVVQQOTGIGIPQHKVALMPQERRRSDRR 431
DB 956 LVLAGTEARDSTVKGKGMONLQDSATKVN-----SGE--YITLGRSKDEFQALARAE 1007
QY 432 DQO--GSVASTHSDSDSEVVNVAEYGGGARNLSARHQPPEHIEVAADPGVSVTONFS 489
DB 1008 DLQVAGGTAIYAGTAD-----ASVSGATGSLSLMTPRDNV-----TPVKLE 1050
-----
OY 490 GSGPVTGRLTGPQGIQSTYALL--ANSGLRLMGGLTSG-----GET 532
DB 1051 GAVRTDTSATLTGLNGVDVTTLADLTAAASRGSVWLNNSNCAGTSNCEYRVNSLLNDGDV 1110
OY 533 AVSSVNAAP 541
DB 1111 YLSAQTAAP 1119
-----
RESULT 7
ID PGCA_RAT STANDARD; PRT; 2124 AA.
AC P07897;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE AGGREGAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE
DE PROTEIN) (CSPCP).
DE AGCL OR AGC.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=88087070; PubMed=3693370;
RA Doegge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RT "Complete primary structure of the rat cartilage proteoglycan core
RT protein deduced from cDNA clones.";
RL J. Biol. Chem. 262:17757-17767(1987).
RN [2]
RN REVISION TO 698.
RA Doegge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RL J. Biol. Chem. 263:10040-10040(1988).
RN [3]
RN SEQUENCE OF 1856-2124 FROM N.A.
RX MEDLINE=86250698; PubMed=2424893;
RA Doegge K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;
RT "Partial cDNA sequence encoding a globular domain at the C terminus
RT of the rat cartilage proteoglycan.";
RL J. Biol. Chem. 261:8108-8111(1986).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CHAINS. N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M13518; AAA41836.1;
-----
```

```
DR EMBL; J03485; AAA21000.1; ALT_SEQ.
DR PIR; A23835; A23835.
DR PIR; A28452; A28452.
DR HSP; P98066; IISG.
DR InterPro; IPR000436; -.
DR InterPro; IPR000495; -.
DR InterPro; IPR000538; -.
DR InterPro; IPR001304; -.
DR Pfam; PF00193; Xlink; 4.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS01241; LINK; 4.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
KW Repeat; Immunoglobulin domain.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2124 AGGREGAN CORE PROTEIN.
FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 170 247 LINK 1.
FT DOMAIN 268 349 LINK 2.
FT DOMAIN 504 581 LINK 3.
FT DOMAIN 602 683 LINK 4.
FT DOMAIN 1910 2036 C-TYPE LECTIN.
FT DOMAIN 2040 2098 SUSHI.
FT DOMAIN 48 140 G1-A.
FT DOMAIN 152 247 G1-B.
FT DOMAIN 253 349 G1-B'.
FT DOMAIN 486 580 G2-B.
FT DOMAIN 587 682 G2-B'.
FT DOMAIN 685 798 KS.
FT DOMAIN 801 1226 CS-1.
FT DOMAIN 1227 1909 CS-2.
FT DOMAIN 1910 2124 G3.
FT DISULFID 51 133 BY SIMILARITY.
FT DISULFID 175 246 BY SIMILARITY.
FT DISULFID 199 220 BY SIMILARITY.
FT DISULFID 273 348 BY SIMILARITY.
FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 509 580 BY SIMILARITY.
FT DISULFID 533 554 BY SIMILARITY.
FT DISULFID 607 682 BY SIMILARITY.
FT DISULFID 631 652 BY SIMILARITY.
FT DISULFID 1914 1925 BY SIMILARITY.
FT DISULFID 1942 2034 BY SIMILARITY.
FT DISULFID 2010 2026 BY SIMILARITY.
FT DISULFID 2041 2084 BY SIMILARITY.
FT DISULFID 2070 2097 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1842 1842 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2124 AA; 221117 MW; E30BBE61593A34B1 CRC64;

Query Match 4.8%; Score 135.5; DB 1; Length 2124;
Best Local Similarity 20.7%; Pred. No. 4.2;
Matches 138; Conservative 77; Mismatches 204; Indels 247; Gaps 33;

QY 17 PPAPPLPSQTD----GAARGGTGHL-----ISTGALGSRSLFSPLRNMSVDSR 64
DB 880 PTVDRLPSGSGESLEGSASAGTDLGSLPSGGEITTSAG-----TE 922
QY 65 DIPGLPTNPSRLAAATSETCLLGGFEVL-HDKGPDILNTQIGPSAFRVEQADGTHAI 123
DB 923 EISGLPSGGDDLETSTSG---IDGASVLTPTGRGGL-----ETSASGVEDLS 965
QY 124 GKNGLEVSVTLSPOWSSLSQSDT--EGKNRFVETGGRG-----GSGHPMVTVASDIA 175
```

```
Db 966 GLPSGEGSETST---SGIEDISVLPTGESPETASGVGLDGLSPGSGESLETASGVE 1021
QY 176 EARTILAKLPDNGHGRQPKVDYRSCVGVSGASGIDDDGVVSET---HTSTNSSVRS DP 232
Db 1022 DV-TQL-----PTERGGLT-----SASGIEDITVLPTGRENLETASGVE--- 1061
QY 233 KFWVSVAIAAGLAGL--AATGI-----AQAALALT-----PEPDDPTTDP- 271
Db 1062 ----DVSGLPSCKEGLETASGIEDISVPTAEAGLETASGGYVSGIPSGEDGTST 1117
QY 272 -----DQANAAS-----ATKDLQTOFAKNPENQKNVIDANGNAIPSGE- 312
Db 1118 GVEGVSLPSGCGELETASGVVEDLGLPTRDSL-----ETSASGVDTG--YPSGRE 1167
QY 313 -----LXDDIVEIAQQAQAEVAAQQAQRYEDQHARQBELQLSSG- 363
Db 1168 DTETSPVGVGDDL-----SGLPSGQEGLETASGAEDLGGPSGKEDLVGSASGA 1217
QY 364 -----IGYGLSSALIVAGGICAGVTTALHRR 389
Db 1218 LDFGLPSGTLCGGQTPEASGLPSGFSGEYSGVDIGSGPSSGLPDFSGLPSGFPVSLVD 1277
QY 390 NQPAEQTTTTTHVVQQQT-----GGIPQHKVAMPQERRRFRDRRSGQSVAST 440
Db 1278 STLVEVITATTASELEGRGTISVSGSGEESGPPSEL-----DSSADISGLPSGT 1327
QY 441 HWSDSSEVVNPNYAEVGGGARNLSAHOPEHYDEVAADPGYVIONFSGSGPVTGLIC 500
Db 1328 ELSGQTSGSLDVSGETSG-----FFDVSGQFPFGS-----SGTGGTSGIPE 1368
QY 501 TPGQGIQS--TVALLANS--GLRLGMG---GLTS--GGETA--VSSVNA 540
Db 1369 VSGQAVRSPDTTEISELSGLSSGQPDVSGEGSGLIFGSGSGSITSVSGTSGISDL 1428
QY 541 PTPGPV 546
Db 1429 PSGFPV 1434

RESULT 8
APMO_PIG STANDARD; PRT; 1150 AA.
AC P12021;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE APOMUCIN (MUCIN CORE PROTEIN) (FRAGMENT).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Submaxillary gland;
RA MEDLINE=91236743; PubMed=2033060;
RX Eckhardt A.E., Timpte C.S., Abernethy J.L., Zhao Y., Hill R.L.;
RT "Porcine submaxillary mucin contains a cysteine-rich,
RT carboxyl-terminal domain in addition to a highly repetitive,
RT glycosylated domain.";
RL J. Biol. Chem. 266:9678-9686(1991).
RN [2]
RP SEQUENCE OF 1-503 FROM N.A.
RC TISSUE-Submaxillary gland;
RX MEDLINE=88087170; PubMed=2826455;
RA Timpte C.S., Eckhardt A.E., Abernethy J.L., Hill R.L.;
RT "Porcine submaxillary gland apomucin contains tandemly repeated,
RT identical sequences of 81 residues.";
RL J. Biol. Chem. 263:1081-1088(1988).
RN [3]
RP SEQUENCE OF 45-80.
RC TISSUE-Submaxillary gland;
RX MEDLINE=87280230; PubMed=3611111;
RA Eckhardt A.E., Timpte C.S., Abernethy J.L., Toumadje A.,
```

| | | | | |
|--------------|------|------------|------------------------|---------------|
| Carbohydrate | 73 | 73 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 74 | 74 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 76 | 76 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 77 | 77 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 81 | 81 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 83 | 83 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 87 | 87 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 91 | 91 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 93 | 93 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 94 | 94 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 96 | 96 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 98 | 98 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 101 | 101 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 103 | 103 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 104 | 104 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 106 | 106 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 107 | 107 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 110 | 110 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 111 | 111 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 117 | 117 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 123 | 123 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 124 | 124 | O-LINKED (GALNAC. . .) | (MUCIN TYPE). |
| Carbohydrate | 418 | 418 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| Carbohydrate | 547 | 547 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| Carbohydrate | 917 | 917 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| Carbohydrate | 985 | 985 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| Carbohydrate | 1002 | 1002 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| Carbohydrate | 1068 | 1068 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| SEQUENCE | 1150 | AA; 109615 | MM; 3CB68B5D29DD7F5A | CRC64; |

Query Match 4.8%; Score 134.5; DB 1; Length 1150;

Best Local Similarity 22.1%; Pred. No. 2.2;

Matches 126; Conservative 60; Mismatches 260; Indels 125; Gaps

| | | |
|-----|---|-----|
| 24 | SOTDGAARGGTGHLISSTGALGS--RSLFSPLRNSMADSVDSRDIPGLTPNPSRLAAATS | 81 |
| 91 | SGTTGTVSGAGSTGSSGSPGATGASIQGP-ETSRISVAGSSGAPVSSGASQ-AACTS | 148 |
| 82 | ETCLLGGFEVLHDKPLDLNTIQIGPSAFRVEVQADGTHAAGEKNGLEVSVTLSQPEWS | 141 |
| 149 | -----GAGPGTTASSGVGTETARPSVAGSGTTGTVSGAGSGTGGSSGSPGATG | 196 |
| 142 | SLQSIDTEGKNRFVTGRRGG-----SGHPWTVVASDIAEARTRILAKLDPD | 188 |
| 197 | A--SIGQPETSRISVAGSSGAPVSSGASQAAGTSGAGPGTTASSGVGTET-----ARPS | 249 |
| 189 | NHGGRQPKDVTFRVGVGSGASGIDGGV---VSEHTSTTNSVRSDPKFWVSVGAIAAGL | 245 |
| 250 | VAGSGTGTVSGAGSTGSSGSPGATGASIQGPETSRISVAGSSGAPVSSGASQAAGT | 309 |
| 246 | AG-----LAATGIAQALATPEPDPTTDPDQAANAASATKQOLTOEAFKNFENOKV | 299 |
| 310 | SGAGPGTTASSGVGTETARPSVAGSGTCTVSGAGSGTGGSSGSPGATGASIQGPETSR | 369 |
| 300 | NI--DANGNAIPSGELXDDIVEIOAQAKEAGEVARQQAQAVESNAQAQORYEDQHARRQBE | 357 |
| 370 | SVAGSSGAPVSSG-----ASQA--AGTSEATTSIEGAG----- | 401 |
| 358 | LOLSSGIGVGLLSALIVAG-----GIGAGVTALHRRN-QP-----ABQTTT-----TT | 401 |
| 402 | ---TSGVGP-KTEATTPGENETTRVGATGTTGIVSRKLTLEPGSYNTEATTSGRSGTT | 457 |
| 402 | HTVVQOQTGGIPQHKVQALMPQERRRSDRDSOGSVASTHWSDSSESVNPAEUVGARN | 461 |
| 458 | HT-----DLPQGTITVLPGFSH--SSGSKPGSVVTPGSPESGSETGTSGEFSTVI | 508 |
| 462 | SLSAHQPEEHIYDEVAADPGYSVIQNFSGSGP-VTGRGLIGTP-GQGIGQISTALLANS | 519 |
| 509 | SGSH-TEATTFIGGSGPG-----TGSRPGTTGELSGCTTIASGNATTEA--TTSTET | 558 |
| 520 | RLG-----MGULTSGGETAVSVNAP | 541 |


```

Schizophyllaceae: Schizophyllum.
[1]
NCBI_TaxID=5334;
RP SEQUENCE FROM N.A.
RC STRAIN=UVM 4-40;
RX MEDLINE=92357793; PubMed=1353886;
RA Stankis M.M., Specht C.A., Yang H., Giasson L., Ullrich R.C.,
RA Novotny C.P.;
RT "The A alpha mating locus of Schizophyllum commune encodes two
RT dissimilar multiallelic homeodomain proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7169-7173(1992).
CC -!- FUNCTION: SPECIES A-ALPHA-4 MATING-TYPE. MAY REGULATE THE
CC EXPRESSION OF GENES SPECIFIC TO THE HOMOKARYOTIC CELL TYPE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED CONSTITUTIVELY IN HOMOKARYONS.
CC -!- SIMILARITY: BELONGS TO THE TALE/M-ATYP FAMILY OF HOMEBOX
CC PROTEINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M97181; AAB01372.1; -.
DR PIR; D37271; D37271.
DR InterPro; IPR001356; -.
DR PROSITE; PS00027; HOMEBOX_1; FALSE-NEG.
DR PROSITE; PS00071; HOMEBOX_2; FALSE-NEG.
KW Homeobox; DNA-binding; Transcription regulation; Nuclear protein.
FT DNA_BIND 110 182 HOMEBOX (TALE-TYPE).
FT DOMAIN 370 402 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 436 469 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 940 AA; 101856 MW; 4B99CBAEDB39621E CRC64;
-----
Query Match 4.7%; Score 130.5; DB 1; Length 940;
Best Local Similarity 22.5%; Pred. No. 2.9;
Matches 113; Conservative 63; Mismatches 205; Indels 121; Gaps 21;
QY 23 PSQTGARGGTHLISITGALGSLRSLSP-----LRNSMADSVDSR--DIPGLPTN 72
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 470 PVKIAGAKRGNDDEVSPLAKKPR-IFSPVPRPQAIRVSLPSPASSRGSTPTSPVS 528
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 73 PS---RLAATSEFCLLGGPEV-LHDKGDLILNTQGPSAFRVEVQADGTH-AAIGEK 126
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 529 PSPKAKRPAQATS---LLASHPMKKREKLQELRKAGLAPPSAPVLMGPDGVPGLGTVRSR 585
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 127 NGLVSVTLSPQESWSLQSIDTEGKNRNVFTGGRGGSGHPMVTVAASDAEARTILAKLD 186
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 586 SSPSVS---SPSPSVSLPLPSRG---VPSGKIKVTGDPFPWNWDL-EAHT----- 630
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 187 PDNHGGRPKD-----VD-----TRSVGVGSAGIDGCVSVSTHTST 223
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 631 -----QAPRDLTAATKSSAGCVDAVPLPGKRSRLTRSPSSIS-----SACST 675
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 224 TNSVSRDPKFWVSGAIAAGLAGLAATGIAQAALATPEDDPTTDDQAAANAESATK 283
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 676 SSSGSDTDSLSFVT-----SDATDIT-EPDEATTADETTTQSTSASSSR 718
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 284 DLTQEAEPKNPNQKNVID-----ANGNAIPSGEL-----XDDIVEQIA 322
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 719 DTSQOKRMPPUSIDPRDPALWSYDLSPPADGRHLHPSDGLRPSAFVPTKLDVRVANLA 778
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 323 QOAKEAGEVARQOAEQAQRYEDQHARRQEELQLSSGIGYGLSSALIIVAG---GIG 379
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 779 QNPARHWSAKSPRTFASHAAAPVSYHHATGSIASPAQVAFEGQLTSLVATGQKAGNA 838
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 380 AGVTTALHRRNOPASQTTTTTHTVVQQTGGIPI-----HKVALMPQRRRRFRSDRSQ 434
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 839 RRRRTTPQVRVTPKAQETSEPSLVDGILSSGLADVCREAPKAPKAKPNDRRLERRRR 898
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```
FT CARBOHYD 687 687 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 723 723 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 820 820 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 886 886 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 918 918 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 973 973 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1068 1068 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1260 AA; 132641 MW; 763D1063A2354C24 CRC64;

Query Match 4.6%; Score 130; DB 1; Length 1260;
Best Local Similarity 20.6%; Pred. No. 4.4;
Matches 73; Conservative 48; Mismatches 167; Indels 66; Gaps 12;

QY 16 IPAPPLP---SOTDGAARCGTGHLLSS-----TCALGSRSLFSPRLNSMADSVDSRD 65
DB 895 VPTASTMSDSLSTDSGTISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTLQLSS 954
QY 66 IGLPTNPSRLAAATSETCLLGGFEVLHDKGLDILNTQIGPSAFRV--EQVADGTHAAI 123
DB 955 IPSVSESEKVTTSNGDNOSGTHDSQSTSTEIEIVT-----STKVLPPVVSNTDLTS 1010
QY 124 GEKNGLSVTLSPQEWSSLSQSIDTE---CKNRFVFTGGRGGSGHPMVTVA-SDIAEART 179
DB 1011 EPTNREQPTTLSTTSNISITEDITTSQPTGDN-----GDNTSNTNPTVATSTLASAS- 1064
QY 180 RILAKLPDNNHGROPKDVTNRSGVSGASGIDGCVVSETHSTTSSVRSRDKFVWSVG 239
DB 1065 -----EEDNKGSGH-----ESASTSLKPSMGENSEGLTSTEIEATTTSTPAFSPVSSG 1114
QY 240 AIAAGLAGLAATGAALATPPDD-----PTTTPDQAAANAASATKDOLTQFAEKNPE 295
DB 1115 T-----DVTETPTDREQPTTLSTTSKTNSESVATQTATNENGGRKSPS 1157
QY 296 NQ-----KVNIDANGNAIPSGELXDDIVEIOIAQAKAEAGEVARQAVESN 340
DB 1158 TDLTSSLTGTSTASTSANSSELVTSGVGGAVASANDQSHSTSVTNSNIVSN 1211
```

```
RESULT 14
MAM_DROME STANDARD; PRT; 1596 AA.
AC P21519;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE NEUROGENIC PROTEIN MASTERMIND.
GN MAM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=91065516; PubMed=1701150;
RA Smoller D., Friedel C., Schmid A., Bettler D., Lam L.,
RA Yedvobnick B.;
RT "The Drosophila neurogenic locus mastermind encodes a nuclear protein
RT unusually rich in amino acid homopolymers.";
RL Genes Dev. 4:1688-1700(1990).
CC -1- FUNCTION: MAY HAVE A REGULATORY FUNCTION POSSIBLY IN ASSOCIATION
CC WITH THE N GENE PRODUCT.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DEVELOPMENTAL STAGE: DURING EARLY NEUROGENESIS MAM PRODUCTS ARE
CC UBICUOUSLY LOCATED. DURING LATER STAGES THEY ACCUMULATE IN THE
CC CENTRAL NERVOUS SYSTEM.
CC -1- MISCELLANEOUS: THE PROTEIN HAS MANY AA HOMOPOLYMERIC DOMAINS: 21
CC POLY-GLN RUNS (FROM 5 TO 16 AA IN LENGTH), 4 POLY-GLY (6 TO 10
CC AA), 3 POLY-ASN (3 X 5 AA), 1 POLY-ALA (10 AA) AND 1 POLY-THR (5
CC AA) RUNS.
```

```
CC -1- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
CC YEAST AND MAMMALIAN REGULATORY PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X54251; CAA38152.1;
CC PIR: A33106; A33106.
CC PIR: A36391; A36391.
CC FlyBase: FBgn0002643; man.
KW Neurogenesis; Nuclear protein; Repeat.
FT DOMAIN 20 84
FT DOMAIN 127 190 ARG/LYS-RICH (BASIC).
FT DOMAIN 196 219 GLN-RICH.
FT DOMAIN 259 304 ASN-RICH.
FT DOMAIN 355 388 GLY/ASN-RICH.
FT DOMAIN 392 406 GLN-RICH.
FT DOMAIN 407 440 GLY-RICH.
FT DOMAIN 651 671 GLN-RICH.
FT DOMAIN 700 714 GLN-RICH.
FT DOMAIN 759 816 5 X 2 AA TANDEM REPEATS OF G-V.
FT DOMAIN 987 996
FT DOMAIN 1060 1079 ALA-RICH.
FT DOMAIN 1092 1107 8 X 2 AA TANDEM REPEATS OF V-G.
FT DOMAIN 1237 1252 7 X 2 AA TANDEM REPEATS OF G-V.
FT DOMAIN 1492 1496 POLY-THR.
FT DOMAIN 1559 1592 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 1596 AA; 167717 MW; B944D86EF359D605 CRC64;

Query Match 4.6%; Score 130; DB 1; Length 1596;
Best Local Similarity 20.6%; Pred. No. 5.9;
Matches 68; Conservative 50; Mismatches 140; Indels 72; Gaps 12;

QY 262 EPDDPTTTPDQAAANAASATKDOLTQFAEKNPENQKVNID-----ANGNAIPSGE---L 313
DB 873 QQNNPTGPGGNTPNAPQOQOQOQSTTTTLQMKQTQLHISQOQGGAGIQVSAQHLHL 932
QY 314 XDDIVEIOIAQAKAEAGEVARQAVESNAQAQRYEDQHAREEELQ-----LSSGIGVGL 368
DB 933 SGDMKSNVVAAGQVFFSOQAQOQOQOQOQPGGTNGPNPQOQOQOQPHGGNAGGVGVG- 991
QY 369 SSALIVAGGIGAGVTTALHRRNQPAEQTTT-----TTTHTVVVQOQTGGIPQ 414
DB 992 -----VGVGVGNGGPNFCOQOQOQPNQNSNANVPDGFSLSQSQSMNPNQOQOQAAAAQ 1046
QY 415 HKVALMFOERRRFSRRDSQGSVASTHWSDSSEVVPYAEVGGARNSLSAHOPEEHYD 474
DB 1047 QQV--QPNMRQRTQAQAAAAAQAQAAAAAASGP-----NVPLMQQPVQV---- 1092
QY 475 EVAADPGVSVIQFSGSGPVTVGLTGPQGIQSTYALLANSGLRLGMLTSGGETAV 534
DB 1093 GVGVGVGVG-----GVGVGVGVGVGVGGGPPNNGA--MNQMGPGMGPGMGPGMPNP 1146
QY 535 SSVN---AAPT-----PGP 545
DB 1147 MQMNPNAAGPTAQOQMMGSGAGGPGQVPGP 1176

RESULT 15
NOTC_DROME STANDARD; PRT; 2703 AA.
AC P07207; P04154;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.
GN N.
```


OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86079539; PubMed=3935325;
 RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
 RT "Nucleotide sequence from the neurogenic locus notch implies a gene
 RT product that shares homology with proteins containing EGF-like
 RT repeats.";
 RL Cell 43:567-581(1985).
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=87064624; PubMed=3097517;
 RA Kidd S., Kelley M.R., Young M.W.;
 RT "Sequence of the notch locus of Drosophila melanogaster: relationship
 RT of the encoded protein to mammalian clotting and growth factors.";
 RL Mol. Cell. Biol. 6:3094-3108(1986).
 RP SEQUENCE OF 2505-2611 FROM N.A.
 RX MEDLINE=85093929; PubMed=2981631;
 RA Wharton K.A., Yedvobnick B., Finnerty V.G., Artavanis-Tsakonas S.;
 RT "opa: a novel family of transcribed repeats shared by the Notch locus
 RT and other developmentally regulated loci in D. melanogaster.";
 RL Cell 40:55-62(1985).
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE=87257846; PubMed=3037327;
 RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;
 RT "Restriction of P-element insertions at the Notch locus of Drosophila
 RT melanogaster.";
 RL Mol. Cell. Biol. 7:1545-1548(1987).
 RP REVIEW.
 RA Harris W.A.;
 RT "Many cell types specified by Notch function.";
 RL Curr. Biol. 1:120-122(1991).
 CC -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
 CC ECTODERM.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
 CC THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS
 CC DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE
 CC NEUROGENIC GENES.
 CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M16152; AAB59220.1; -
 CC EMBL; M16153; AAB59220.1; JOINED.
 CC EMBL; M16149; AAB59220.1; JOINED.
 CC EMBL; M16150; AAB59220.1; JOINED.
 CC EMBL; M16151; AAB59220.1; JOINED.
 CC EMBL; K03508; AAA28725.1; -
 CC EMBL; M13689; AAA28725.1; JOINED.
 CC EMBL; K03507; AAA28725.1; JOINED.
 CC EMBL; M12175; AAA74496.1; -
 CC EMBL; M16025; AAA28726.1; -
 CC PIR; A24420; A24420.
 CC PIR; A24768; A24768.
 CC PIR; A05267; A05267.

DR HSSP; P00740; IIXA.
 DR FlyBase; FBgn0004647; N.
 DR InterPro; IPR000152; -
 DR InterPro; IPR000561; -
 DR InterPro; IPR000800; -
 DR InterPro; IPR001438; -
 DR InterPro; IPR001881; -
 DR InterPro; IPR002110; -
 DR Pfam; PF00008; EGF; 36.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00066; notch; 3.
 DR PRINTS; PR00010; EGFBL00D.
 DR PROSITE; PS50088; ANK_REPEAT; 5.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 28.
 DR PROSITE; PS01187; EGF_CA; 22.
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 44
 FT CHAIN 45 2703
 FT DOMAIN 45 1745
 FT TRANSMEM 1746 1766
 FT DOMAIN 1767 2703
 FT DOMAIN 58 95
 FT DOMAIN 96 136
 FT DOMAIN 139 176
 FT DOMAIN 177 215
 FT DOMAIN 217 253
 FT DOMAIN 255 291
 FT DOMAIN 293 329
 FT DOMAIN 331 370
 FT DOMAIN 372 408
 FT DOMAIN 409 447
 FT DOMAIN 449 486
 FT DOMAIN 488 524
 FT DOMAIN 526 562
 FT DOMAIN 564 600
 FT DOMAIN 602 637
 FT DOMAIN 639 675
 FT DOMAIN 677 713
 FT DOMAIN 715 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 865
 FT DOMAIN 867 905
 FT DOMAIN 907 944
 FT DOMAIN 946 982
 FT DOMAIN 984 1020
 FT DOMAIN 1022 1058
 FT DOMAIN 1060 1096
 FT DOMAIN 1098 1134
 FT DOMAIN 1136 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1257
 FT DOMAIN 1259 1295
 FT DOMAIN 1297 1335
 FT DOMAIN 1337 1373
 FT DOMAIN 1375 1412
 FT DOMAIN 1415 1451
 FT REPEAT 1475 1513
 FT REPEAT 1514 1553
 FT REPEAT 1554 1593
 FT REPEAT 1591 1627
 FT REPEAT 1628 1664
 FT REPEAT 1665 1701
 FT REPEAT 1702 1738
 FT REPEAT 1739 1775
 FT REPEAT 1776 1812
 FT REPEAT 1813 1849
 FT REPEAT 1850 1886
 FT REPEAT 1887 1923
 FT REPEAT 1924 1960
 FT REPEAT 1961 1997
 FT REPEAT 1998 2034
 FT REPEAT 2035 2071
 FT REPEAT 2072 2108
 FT REPEAT 2109 2145
 FT REPEAT 2146 2182
 FT REPEAT 2183 2219
 FT REPEAT 2220 2256
 FT DISULFID 62 73
 FT DISULFID 67 83

POLY-GLN (OPA-REPEAT).
 BY SIMILARITY.
 BY SIMILARITY.

| | | | | |
|----|------|---|--|------|
| Db | 2424 | QSSSLASPHAYLGSPSPAKSPSLPTSPTHIQAMRHATQ----- | QKQFGGSNLNLL | 2476 |
| QY | 292 | -----KNPENOKVN-----IDANGN----- | AIPSGELXDDIVEQIA | 322 |
| Db | 2477 | GGANGGVGGGGGGGVGGQGPQNSPYSGLIISPTGSDMGIMLAPQSSKNSAIMQTIS | 2536 | |
| QY | 323 | QOAKEAGEVARQOAVESNAQOORYEDQHARQEELQSSGGYGLSALIVAGGICAGV | 382 | |
| Db | 2537 | PQQQQ----- | QOQQQQQQOHOQQQQQQOQQQQQQOQLGGLEFG-SAGLDLNGFCGSPD | 2589 |
| QY | 393 | TTALHRRNQPAEQT----- | TTTTTH-----TVVQOQFGG-IPQHKVALLM | 420 |
| Db | 2590 | SFHSQGNPPFSIOSSSGSGSPSTMULSPSQNOQAIFYQYLTPSSQHSGGHTPQHLVQTL | 2649 | |
| QY | 421 | PQERRRSDRRDSQGSVASTHSDSSSEVYVNPYAE-- | VGGARNSL-----SAHQPEEHI | 472 |
| Db | 2650 | -----DSYPTSPSPGCHWSSSSPSRNSDWSQGVQSPAAANNLYISGGHOANKGSEAI | 2701 | |
| QY | 473 | Y | 473 | |
| Db | 2702 | Y | 2702 | |

Search completed: September 27, 2001, 14:22:50
Job time: 146 sec

```

Query Match      4.6%; Score 130; DB 1; Length 2703;
Best Local Similarity 20.0%; Pred. No. 11;
Matches 120; Conservative 81; Mismatches 194; Indels 206; Gaps 28;

Qy   4 GNLGNNVNGNHLIPAPPPLPSQT---DGAAR-----GGTGHLLISTGALGSRSLFSPLRN 55
    ||||| : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
Db 2177 GGNNGNNGN----ASGKQSNTAKKAARKAKLEGSPDNGLDATSLRRK---ASSKK 2229
                                     : ||| | : ||| | : ||| | : ||| | :

Qy   56 SMADSVDSRDIPTNPSPRLAAATSETCLLGGFFVLHDKGPDLILNTQIGSAPRVEVQ 115
    ||||| : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
Db 2230 TSAASKKANLNL--NPGQLT-----GGVGVPVCPPTN-----SAQAANA 2270
                                     : ||| | : ||| | : ||| | : ||| | :

Qy   116 ADGTHAAIG-EKNGLLEVSVTLSPQWSSLQSIDTEGKNRFVTGGRGSGHPMWTVASDI 174
    ||| : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
Db 2271 AAAVAAMSHLEGPSVGV-----GMGGNLPSPYDTSSMY 2305
                                     : ||| | : ||| | : ||| | : ||| | :

Qy   175 ABARTRIALLDPDNHGGRQPKVD-----TRSVGSGSAGIDDGWVSEHTSTTNSSVR 229
    ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
Db 2306 SNAMAAPLANGNP-NTGAKQPSYEDCIKNAGSMOSLOCNGL-DMIKLNDYAIYMGSFPQ 2363
                                     : ||| | : ||| | : ||| | : ||| | :

Qy   230 SD-----PKEFWISGATAAGLAGL-----AATG----- 252
    : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
Db 2364 QELLNGOGLMGNGNRQNVGPGVLPFGCGMGGLSGAGNCSHBEGGLSPPYSNOSPPHS 2423
    : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :

Qy   253 IAOALALTP-----EPDDPTTTDPOAANAASEATKDOLTQRAF----- 291
    : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:21:49 ; Search time 38.06 Seconds
(without alignments)
1908.446 Million cell updates/sec

Title: US-09-189-415A-2
Perfect score: 2800
Sequence: 1 MPIGNLGNVNGNHLIPPAP.....GETAVSSVNAAPTPGPVRFV 549

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_unclassified:*
- 13: sp_vertebrate:*
- 14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 2795 | 99.8 | 549 | 2 O50190 | O50190 escherichia |
| 2 | 2676.5 | 95.6 | 550 | 2 O52147 | O52147 escherichia |
| 3 | 2496.5 | 89.2 | 552 | 2 O9KWH9 | O9KWH9 escherichia |
| 4 | 2206 | 78.8 | 551 | 2 O68258 | O68258 escherichia |
| 5 | 2077 | 74.2 | 547 | 2 Q9WKK1 | Q9WKK1 escherichia |
| 6 | 2076 | 74.1 | 547 | 2 Q9ETI1 | Q9ETI1 citrobacter |
| 7 | 1781.5 | 63.6 | 538 | 2 Q47016 | Q47016 escherichia |
| 8 | 1781.5 | 63.6 | 538 | 2 Q47014 | Q47014 escherichia |
| 9 | 1763.5 | 63.0 | 538 | 2 O85508 | O85508 escherichia |
| 10 | 1528.5 | 54.6 | 558 | 2 Q9R396 | Q9R396 escherichia |
| 11 | 1516.5 | 54.2 | 558 | 2 O85506 | O85506 escherichia |
| 12 | 168.5 | 6.0 | 1158 | 5 O97169 | O97169 drosophila |
| 13 | 160.5 | 5.7 | 1818 | 6 Q9N1P0 | Q9N1P0 bos taurus |
| 14 | 159.5 | 5.7 | 1265 | 2 Q9FDA0 | Q9FDA0 xanthomonas |
| 15 | 156.5 | 5.6 | 1323 | 5 Q9NHX4 | Q9NHX4 drosophila |
| 16 | 152 | 5.4 | 1203 | 5 Q9N5K0 | Q9N5K0 caenorhabdi |
| 17 | 151.5 | 5.4 | 2806 | 2 Q9KXA6 | Q9KXA6 escherichia |
| 18 | 151.5 | 5.4 | 2806 | 9 Q9T1K9 | Q9T1K9 bacterioph |
| 19 | 151 | 5.4 | 934 | 5 Q9VN59 | Q9VN59 drosophila |

| | | | | | | |
|----|-------|-----|------|----|--------|--------------------|
| 20 | 150 | 5.4 | 1005 | 2 | O84462 | O84462 chl: |
| 21 | 149.5 | 5.3 | 2806 | 9 | O9XJM1 | O9XJM1 bac |
| 22 | 149 | 5.3 | 832 | 2 | O54356 | O54356 mor |
| 23 | 149 | 5.3 | 2232 | 5 | P91365 | P91365 caenorhabdi |
| 24 | 146 | 5.2 | 705 | 5 | O9GYB2 | O9GYB2 leishmania |
| 25 | 146 | 5.2 | 1829 | 5 | O22248 | O22248 caenorhabdi |
| 26 | 145.5 | 5.2 | 848 | 5 | O9V889 | O9V889 drosophila |
| 27 | 145.5 | 5.2 | 2586 | 5 | O9VTK8 | O9VTK8 drosophila |
| 28 | 144.5 | 5.2 | 1296 | 5 | O96506 | O96506 drosophila |
| 29 | 144 | 5.1 | 1589 | 6 | O62672 | O62672 bos taurus |
| 30 | 143.5 | 5.1 | 1778 | 5 | O9NE65 | O9NE65 leishmania |
| 31 | 141.5 | 5.1 | 647 | 3 | O14273 | O14273 schizosacch |
| 32 | 140.5 | 5.0 | 929 | 5 | O9N993 | O9N993 leishmania |
| 33 | 140.5 | 5.0 | 1472 | 5 | O9V358 | O9V358 drosophila |
| 34 | 139 | 5.0 | 1064 | 4 | O9HAN2 | O9HAN2 homo sapien |
| 35 | 139 | 5.0 | 1860 | 5 | O9U129 | O9U129 leishmania |
| 36 | 138.5 | 4.9 | 2353 | 2 | P71401 | P71401 haemophilus |
| 37 | 138 | 4.9 | 1455 | 5 | O9VP05 | O9VP05 drosophila |
| 38 | 138 | 4.9 | 2857 | 5 | O9VXZ1 | O9VXZ1 drosophila |
| 39 | 137 | 4.9 | 1066 | 11 | O9ERC7 | O9ERC7 mus musculu |
| 40 | 137 | 4.9 | 1475 | 5 | O9GYW7 | O9GYW7 drosophila |
| 41 | 137 | 4.9 | 2468 | 2 | O9I2M3 | O9I2M3 pseudomonas |
| 42 | 136.5 | 4.9 | 770 | 3 | O9P3I9 | O9P3I9 neurospora |
| 43 | 136.5 | 4.9 | 1829 | 2 | O86808 | O86808 streptomyce |
| 44 | 136 | 4.9 | 540 | 9 | O9G0H8 | O9G0H8 roseophage |
| 45 | 136 | 4.9 | 674 | 2 | O9KQ73 | O9KQ73 vibrio chol |

ALIGNMENTS

| | |
|---|---|
| RESULT 1 | |
| O50190 | |
| ID O50190 | PRELIMINARY; PRT; 549 AA. |
| AC O50190; | |
| DT 01-JUN-1998 | (TREMBLrel. 06, Created) |
| DT 01-JUN-1998 | (TREMBLrel. 06, Last sequence update) |
| DT 01-MAR-2001 | (TREMBLrel. 16, Last annotation update) |
| DE TRANSLOCATED INTIMIN RECEPTOR. | |
| GN TIR. | |
| OS Escherichia coli. | |
| OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; | |
| OC Escherichia. | |
| OX NCBI_TaxID=562; | |
| RN [1] | |
| RP SEQUENCE FROM N.A. | |
| RC STRAIN=E2348/69; | |
| RX MEDLINE=98050926; PubMed=9390560; | |
| RA Kenny B., Devlinney R., Stein M., Reinscheid D.J., Frey E.A., | |
| RA Finlay B.B.; | |
| RT "Enteropathogenic E. coli (EPEC) transfers its receptor for intimate | |
| RT adherence into mammalian cells." | |
| RL Cell 91:511-520(1997). | |
| DR EMBL; AF013122; AAB88410.1; - | |
| DR InterPro; IPR003536; - | |
| DR PRINTS; PR01370; TRNSINTMINR. | |
| SQ SEQUENCE 549 AA; 56843 MW; 40C8B8B234409A08 CRC64; | |

Query Match 99.8%; Score 2795; DB 2; Length 549;
Best Local Similarity 99.6%; Pred. No. 6.4e-155;
Matches 547; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

| | | |
|--------|---|-----|
| QY 1 | MPIGNLGNVNGNHLIPPAPLPSTQDGAARGTGHLISSTGALGSRSLFSPLRNSMAAS | 60 |
| DB 1 | MPIGNLGNVNGNHLIPPAPLPSTQDGAARGTGHLISSTGALGSRSLFSPLRNSMAAS | 60 |
| QY 61 | VPSRDIPLGPTNPRLAAATSETCLLGFEVLHDKGPLDLINTQIGPSAFRVEQADGTH | 120 |
| DB 61 | VPSRDIPLGPTNPRLAAATSETCLLGFEVLHDKGPLDLINTQIGPSAFRVEQADGTH | 120 |
| QY 121 | AAIGKNGLEVSTLSPQEWSSLSQSDTECKNRFVFTGGGSGHPMTVVASDIAEARTR | 180 |
| | | |

Db 121 AAIGKNGLEVSVTLSPQEWSSLSQSIDTEGKNRFVFTGGRGSGHPMTVTASDIAEARTK 180
Qy 181 ILAKLDPDNHGGQPKDVTDRSVGVGSASGIDGCVVSETHSTTNSVSRSDPKFWVSVGA 240
Db 181 ILAKLDPDNHGGQPKDVTDRSVGVGSASGIDGCVVSETHSTTNSVSRSDPKFWVSVGA 240
Qy 241 IAAGLAGLAATGIAQALALTPEDDPTTDPDQAANAESATKDQLTQEAFAKPNQKVN 300
Db 241 IAAGLAGLAATGIAQALALTPEDDPTTDPDQAANAESATKDQLTQEAFAKPNQKVN 300
Qy 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQAQAVESNAQAQORYEDQHARRQELQL 360
Db 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQAQAVESNAQAQORYEDQHARRQELQL 360
Qy 361 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQAQAVESNAQAQORYEDQHARRQELQL 360
Db 361 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQAQAVESNAQAQORYEDQHARRQELQL 360
Qy 420 SSGIGYGLSSALIVAGGIGAGVTTLHRRNOPAEQTTTTHTVVQOQTGGIPQHKVAM 420
Db 420 SSGIGYGLSSALIVAGGIGAGVTTLHRRNOPAEQTTTTHTVVQOQTGGIPQHKVAM 420
Qy 480 POERRRSDRDSQGSVASTHWSDSSEVNPYAEVGGARNLSAHOPEEHYDEVAADP 480
Db 480 POERRRSDRDSQGSVASTHWSDSSEVNPYAEVGGARNLSAHOPEEHYDEVAADP 480
Qy 540 GYSVTQNFSGSPVTRGLIGTPGQGIQSTYALLANSGLRLGMLTSGGETAVSSVNA 540
Db 540 GYSVTQNFSGSPVTRGLIGTPGQGIQSTYALLANSGLRLGMLTSGGETAVSSVNA 540
Qy 541 PTPGPVRFV 549
Db 541 PTPGPVRFV 549

RESULT 2
O52147 PRELIMINARY; PRT; 550 AA.
ID O52147
AC O52147;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.;
RL Mol. Microbiol. 0:0-0(1998).
DR EMBL; AF022236; AAC38390.1; -.
DR InterPro; IPR003536; -.
DR PRINTS; PR01370; TRNSINTIMINR.
SQ SEQUENCE 550 AA; 56509 MW; 19DD08A9BE9251CB CRC64;

Query Match 95.6%; Score 2676.5; DB 2; Length 550;
Best Local Similarity 96.0%; Pred. No. 5e-148;
Matches 531; Conservative 11; Mismatches 11; Indels 7; Gaps 2;

Qy 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Db 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Qy 61 VDSRDIPGLTPNPSRLAAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEVQADGTH 120
Db 61 VDSRDIPGLTPNPSRLAAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEVQADGTH 120
Qy 121 AATGEKNGLEVSVTLSPQEWSSLSQSIDTEGKNRFVFTGGRGSGHPMTVTASDIAEARTK 180
Db 121 AATGEKNGLEVSVTLSPQEWSSLSQSIDTEGKNRFVFTGGRGSGHPMTVTASDIAEARTK 180
Qy 181 ILAKLDPDNHGGQPKDVTDRSVGVGSASGIDGCVVSETHSTTNSVSRSDPKFWVSVGA 240
Db 181 ILAKLDPDNHGGQPKDVTDRSVGVGSASGIDGCVVSETHSTTNSVSRSDPKFWVSVGA 240

Db 181 ILAKLDPDNHGGQPKDVTDRSVGVGSASGIDGCVVSETHSTTNSVSRSDPKFWVSVGA 240
Qy 241 IAAGLAGLAATGIAQALALTPEDDPTTDPDQAANAESATKDQLTQEAFAKPNQKVN 300
Db 241 IAAGLAGLAATGIAQALALTPEDDPTTDPDQAANAESATKDQLTQEAFAKPNQKVN 300
Qy 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQAQAVESNAQAQORYEDQHARRQELQL 360
Db 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQAQAVESNAQAQORYEDQHARRQELQL 360
Qy 361 SSGIGYGLSSALIVAGGIGAGVTTLHRRNOPAEQTTTTHTVVQOQTGGIPQHKVAM 420
Db 361 SSGIGYGLSSALIVAGGIGAGVTTLHRRNOPAEQTTTTHTVVQOQTGGIPQHKVAM 420
Qy 421 POERRRSDRDSQGSVASTHWSDSSEVNPYAEVGGARNLSAHOPEEHYDEV 476
Db 421 POERRRSDRDSQGSVASTHWSDSSEVNPYAEVGGARNLSAHOPEEHYDEV 476
Qy 476 AADPGYVIONFSGSPVTRGLIGTPGQGIQSTYALLANSGLRLGMLTSGGETAVSS 536
Db 476 AADPGYVIONFSGSPVTRGLIGTPGQGIQSTYALLANSGLRLGMLTSGGETAVSS 536
Qy 537 VNAAPTGPVRFV 549
Db 537 VNAAPTGPVRFV 549
Qy 538 VNAAPTGPVRFV 550
Db 538 VNAAPTGPVRFV 550

RESULT 3

Q9KWH9 PRELIMINARY; PRT; 552 AA.
ID Q9KWH9
AC Q9KWH9;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE TIR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-HK01;
RA Abe A., Nagano H.;
RT "Analyses of type III secreted proteins and Tir in enteropathogenic
RT Escherichia coli O157:H45.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB036053; BAA96815.1; -.
DR InterPro; IPR003536; -.
DR PRINTS; PR01370; TRNSINTIMINR.
SQ SEQUENCE 552 AA; 57005 MW; A8D79EE22EE50A4B CRC64;

Query Match 89.2%; Score 2496.5; DB 2; Length 552;
Best Local Similarity 90.3%; Pred. No. 1.4e-137;
Matches 501; Conservative 11; Mismatches 34; Indels 9; Gaps 3;

Qy 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Db 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Qy 61 VDSRDIPGLTPNPSRLAAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEVQADGTH 120
Db 61 VDSRDIPGLTPNPSRLAAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEVQADGTH 120
Qy 121 AATGEKNGLEVSVTLSPQEWSSLSQSIDTEGKNRFVFTGGRGSGHPMTVTASDIAEARTK 180
Db 121 AATGEKNGLEVSVTLSPQEWSSLSQSIDTEGKNRFVFTGGRGSGHPMTVTASDIAEARTK 180
Qy 181 ILAKLDPDNHGGQPKDVTDRSVGVGSASGIDGCVVSETHSTTNSVSRSDPKFWVSVGA 240
Db 181 ILAKLDPDNHGGQPKDVTDRSVGVGSASGIDGCVVSETHSTTNSVSRSDPKFWVSVGA 240

QY 241 IAAGLAGLAATGIAQALALTPEPDDPTTTTDPQAAAEASATKDQLTQAFKNPNQKYN 300
DB 241 IAAGLAGLAATGIAQALALTPEPDDPTTTTDPQAAAEASATKDQLTQAFKNPNQKYN 300
QY 301 IDANGNAIPSGELXDDIVBQIAQAQKEAGEVARQAAVESNAQAQRYEDQHARROEELQ 360
DB 301 IDANGNAIPSGELXDDIVBQIAQAQKEAGEVARQAAVESNAQAQRYEDQHARROEELQ 360
QY 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHVVVQOQTGGIPQHKVALM 420
DB 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHVVVQOQTGGIPQHKVALM 420
QY 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHVVVQOQTGGIPQHKVALM 420
DB 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHVVVQOQTGGIPQHKVALM 420
QY 421 POERRRFS- ---RDSQGSVASTHWSDSSEVWNPYAEVGGARNSLSAHQPEEHYDEV 476
DB 421 POERRRFS- ---RDSQGSVASTHWSDSSEVWNPYAEVGGARNSLSAHQPEEHYDEV 476
QY 477 AADPGYSVLQNSGSGPVTVGRLLIGTPGQIGQISTYALLA-NSGGLRLGMGLTSGGETAV-- 534
DB 477 AADPGYSVLQNSGSGPVTVGRLLIGTPGQIGQISTYALLA-NSGGLRLGMGLTSGGETAV-- 534
QY 535 SSVNAAPTGPVRFV 549
DB 535 SSVNAAPTGPVRFV 549
QY 538 ANANAAPTGPVRFV 552
DB 538 ANANAAPTGPVRFV 552

RESULT 4
O68258 PRELIMINARY; PRT; 551 AA.
ID O68258
AC O68258
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98187918; PubMed=9529069;
RA Voss E., Paton A.W., Manning P.A., Paton J.C.;
RT "Molecular analysis of Shiga toxinigenic Escherichia coli O111:H-
RT proteins which react with sera from patients with hemolytic-uremic
RT syndrome";
RL Infect. Immun. 66:1467-1472(1998).
DR EMBL; AF025311; AAC69249.1; -;
DR InterPro; IPR003536; -;
DR PRINTS; PRO1370; TRNSINTIMINR.
SQ SEQUENCE 551 AA; 56975 MW; 6EC95F76EF0F44CC CRC64;

Query Match 78.8%; Score 2206; DB 2; Length 551;
Best Local Similarity 79.8%; Pred. No. 1.1e-120;
Matches 442; Conservative 39; Mismatches 65; Indels 8; Gaps 3;
QY 1 MPTGNLGNVNNHILPPAPPPLPSQTDGAARGTGHLISSTGALGSRSLFSPLRNSMADS 60
DB 1 MPTGNLGNVNNHILPPAPPPLPSQTDGAARGTGHLISSTGALGSRSLFSPLRNSMADS 60
QY 61 VDSRDIPGLTPNPSRLAAATSETCLLGGFEVLHDKGPLDLINTQIGPSAFRVEVQADGTH 120
DB 61 VDSRDIPGLTPNPSRLAAATSETCLLGGFEVLHDKGPLDLINTQIGPSAFRVEVQADGTH 120
QY 121 AAIKENGLEVSVTLSLPSQWSSLOSIDTEGKRFVFTGGRGSGHPMTVVASDIAEARTR 180
DB 121 AAIKENGLEVSVTLSLPSQWSSLOSIDTEGKRFVFTGGRGSGHPMTVVASDIAEARTR 180
QY 181 ILAKLDPDNHGGQPKDVTNRVSGVSGASGIDGVSSETHSTNTSSSVRSRDPKFWVSVGA 240
DB 181 ILAKLDPDNHGGQPKDVTNRVSGVSGASGIDGVSSETHSTNTSSSVRSRDPKFWVSVGA 240
QY 241 IAAGLAGLAATGIAQALALTPEPDDPTTTTDPQAAAEASATKDQLTQAFKNPNQKYN 300
DB 241 IAAGLAGLAATGIAQALALTPEPDDPTTTTDPQAAAEASATKDQLTQAFKNPNQKYN 300

DB 241 IAAGLAGLAATGIAQALALTPEPDDPTTTTDPQAAAEASATKDQLTQAFKNPNQKYS 300
QY 301 IDANGNAIPSGELXDDIVBQIAQAQKEAGEVARQAAVESNAQAQRYEDQHARROEELQ 360
DB 301 IDANGNAIPSGELXDDIVBQIAQAQKEAGEVARQAAVESNAQAQRYEDQHARROEELQ 360
QY 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHVVVQOQTGGIPQHKVALM 420
DB 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHVVVQOQTGGIPQHKVALM 420
QY 421 POERRRFS- ---RDSQGSVASTHWSDSSEVWNPYAEVGGARNSLSAHQPEEHYDEV 476
DB 421 POERRRFS- ---RDSQGSVASTHWSDSSEVWNPYAEVGGARNSLSAHQPEEHYDEV 476
QY 477 AADPGYSVLQNSGSGPVTVGRLLIGTPGQIGQISTYALLA-NSGGLRLGMGLTSGGETAVS 535
DB 477 AADPGYSVLQNSGSGPVTVGRLLIGTPGQIGQISTYALLA-NSGGLRLGMGLTSGGETAVS 535
QY 536 SSVNAAPTGPVRFV 549
DB 536 SSVNAAPTGPVRFV 549
QY 538 ANANAAPTGPVRFV 551
DB 538 ANANAAPTGPVRFV 551

RESULT 5
O9WXK1 PRELIMINARY; PRT; 547 AA.
ID O9WXK1
AC O9WXK1
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MPEC;
RA Okutani A., Itoh K., Sasakawa C.;
RT "Translocated intimin receptor (Tir) of murine pathogenic Escherichia
RT coli O15a,c:K(B).";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026719; BAA77400.1; -;
DR InterPro; IPR000484; -;
DR InterPro; IPR003536; -;
DR PRINTS; PRO1370; TRNSINTIMINR.
DR ProDom; PD000551; -; 1.
KW Receptor.
SQ SEQUENCE 547 AA; 56283 MW; 02CAC6D625FA6EE1 CRC64;

Query Match 74.2%; Score 2077; DB 2; Length 547;
Best Local Similarity 76.0%; Pred. No. 3.5e-113;
Matches 418; Conservative 44; Mismatches 84; Indels 4; Gaps 3;
QY 1 MPTGNLGNVNNHILPPAPPPLPSQTDGAARGTGHLISSTGALGSRSLFSPLRNSMADS 60
DB 1 MPTGNLGNVNNHILPPAPPPLPSQTDGAARGTGHLISSTGALGSRSLFSPLRNSMADS 60
QY 61 VDSRDIPGLTPNPSRLAAATSETCLLGGFEVLHDKGPLDLINTQIGPSAFRVEVQADGTH 120
DB 61 VDSRDIPGLTPNPSRLAAATSETCLLGGFEVLHDKGPLDLINTQIGPSAFRVEVQADGTH 120
QY 121 AAIKENGLEVSVTLSLPSQWSSLOSIDTEGKRFVFTGGRGSGHPMTVVASDIAEARTR 180
DB 121 AAIKENGLEVSVTLSLPSQWSSLOSIDTEGKRFVFTGGRGSGHPMTVVASDIAEARTR 180
QY 181 ILAKLDPDNHGGQPKDVTNRVSGVSGASGIDGVSSETHSTNTSSSVRSRDPKFWVSVGA 240
DB 181 ILAKLDPDNHGGQPKDVTNRVSGVSGASGIDGVSSETHSTNTSSSVRSRDPKFWVSVGA 240
QY 241 IAAGLAGLAATGIAQALALTPEPDDPTTTTDPQAAAEASATKDQLTQAFKNPNQKYN 300
DB 241 IAAGLAGLAATGIAQALALTPEPDDPTTTTDPQAAAEASATKDQLTQAFKNPNQKYN 300

```
||||| 239 IAAGLAGLAATGIVQAVALTAPDDPTTDPDEAANAEEAATKDQLTKEAFQNPQKNV 298
QY 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQAQVESNAQAQORYEDQHARRQEELOL 360
Db 299 IDELGNAPSGELKDDVVAQIADQAKVAGEAQARQAQVESNAQAQRRHDDQAKRQOELD 358
QY 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNOPAEQTTTTTHVVVQOQTGGIPQHKVALM 420
Db 359 SSGIGYGLSSALIVAGGIGAGVTTALHRRNPTEQ-TIATTHSVIOOQTGGNTRAOGGAD 417
QY 421 PQERRRES--DRRDSQGSVASTHWSDSSEVWNPYAEVGGARNSLSAHOPEEHYDEVAAD 479
Db 418 TTGVENASLTRDSQASVASTQSDTSGDVNPNYABGWSRNPSPLLAPEEPIYDEVAPD 477
QY 480 PGYSVIONFSGSPVTRGLIGTGGQIQSTYALLANSGLRLGMGLTSGGETAVSSVNA 539
Db 478 PNYSVIOHFSGNPNVTGRVLGSPGQIQSTYALLANSGLRLGMGLTSGGESAGSAANA 537
QY 540 APTPGPVRFV 549
Db 538 APTPGVERFV 547

RESULT 7
Q47016 PRELIMINARY; PRT; 538 AA.
ID Q47016;
AC Q47016;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-AUG-1999 (TREMblrel. 11, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-REPEC 83/39;
RA O'Brien R.A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-REPEC 83/39;
RA Krejany E.O.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-RDEC-1;
RA Agin T.S., Boedecker E.C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE OF 172-538 FROM N.A.
RC STRAIN-REPEC RDEC-1;
RA Krejany E.O.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U59504; AAD19750.1; -.
DR EMBL; AF045568; AAC15683.1; -.
DR EMBL; U59503; AAB02941.1; -.
DR InterPro; IPR003536; -.
DR PRINTS; PR01370; TRNSINTMINR.
KW Receptor.
SQ SEQUENCE 538 AA; 55410 MW; 3132A969B7B3D06C CRC64;
```

```
Query Match 63.6%; Score 1781.5; DB 2; Length 538;
Best Local Similarity 65.4%; Pred. No. 5,1e-96;
Matches 367; Conservative 66; Mismatches 93; Indels 35; Gaps 11;

QY 1 MPIGNLGNVNGNHILPPAPPLPSQTDGAARCGTGHLLISSTGALGSRSLFSPLRNSMADS 60
Db 1 MPIGNLGHNSNRALIPAPPLPSQTDGAG-GARNQLNSNGPMGSRLLFTFIRNSVADA 59
```

```
||||| 239 IAAGLAGLAATGIVQAVALTAPDDPTTDPDEAANAEEAATKDQLTKEAFQNPQKNV 298
QY 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQAQVESNAQAQORYEDQHARRQEELOL 360
Db 299 IDELGNAPSGELKDDVVAQIADQAKVAGEAQARQAQVESNAQAQRRHDDQAKRQOELD 358
QY 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNOPAEQTTTTTHVVVQOQTGGIPQHKVALM 420
Db 359 SSGIGYGLSSALIVAGGIGAGVTTALHRRNPTEQ-TIATTHSVIOOQTGGNTRAOGGAD 417
QY 421 PQERRRES--DRRDSQGSVASTHWSDSSEVWNPYAEVGGARNSLSAHOPEEHYDEVAAD 479
Db 418 TTGVENASLTRDSQASVASTQSDTSGDVNPNYABGWSRNPSPLLAPEEPIYDEVAPD 477
QY 480 PGYSVIONFSGSPVTRGLIGTGGQIQSTYALLANSGLRLGMGLTSGGETAVSSVNA 539
Db 478 PNYSVIOHFSGNPNVTGRVLGSPGQIQSTYALLANSGLRLGMGLTSGGESAGSAANA 537
QY 540 APTPGPVRFV 549
Db 538 APTPGVERFV 547

RESULT 6
Q9ETII PRELIMINARY; PRT; 547 AA.
ID Q9ETII;
AC Q9ETII;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR TIR.
OS Citrobacter rodentium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Citrobacter.
OX NCBI_TaxID=67825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1843-73T, AND DBS100;
RX MEDLINE-2055330; PubMed-11101562;
RA Luperchio S.A., Newman J.V., Dangler C.A., Schrenzel M.D.,
RA Brenner D.J., Steigerwalt A.G., Schauer D.B.;
RT "Citrobacter rodentium, the Causative Agent of Transmissible Murine
RT Colonic Hyperplasia, Exhibits Clonality: Synonymy of C. rodentium and
RT Mouse-Pathogenic Escherichia coli.";
RL J. Clin. Microbiol. 38:4343-4350(2000).
DR EMBL; AF301618; AAG40758.1; -.
DR EMBL; AF301617; AAG25642.1; -.
KW Receptor.
SQ SEQUENCE 547 AA; 56270 MW; C88318B301049C37 CRC64;
```

```
Query Match 74.1%; Score 2076; DB 2; Length 547;
Best Local Similarity 76.0%; Pred. No. 4e-113;
Matches 418; Conservative 44; Mismatches 84; Indels 4; Gaps 3;

QY 1 MPIGNLGNVNGNHILPPAPPLPSQTDGAARCGTGHLLISSTGALGSRSLFSPLRNSMADS 60
Db 1 MPIGNLGNVNGNHILPPAPPLPSQTDGATRGSGSLISSTGSLRSLFSPLRNSIVDT 60
QY 61 VDSRDIPGLTPNPSLAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEQADGTH 120
Db 61 VDSRDVPGLEPHLRF--ATSETCLHGGFEVLHDKGPDILNOKIGSSVFRVEQPDGTH 118
QY 121 AAIGKNGLEVSVTLSPQWSSISQSDTEGKNRFVFTGGGSGHPMTVVASDIAEARTR 180
Db 119 AATGVKDGVEVSVTLNSSELQSLSDTETGTRFVTGGGSGHAMVTVASDISQAREK 178
QY 181 ILAKLPDNGHGRQKDVTRSVGVSAGSIDDGVVYVETHSTTTNSSVSRSDPKFWVSVGA 240
Db 179 ILAKLPDNGHGRQKDVTRSVGVSAGSGMDGVVYVETHSTTTSSVSRSDPKFWVSVGA 238
QY 241 IAAGLAGLAATGIAQALATPEPDDPTTDPDQAANAEEAATKDQLTKEAFQNPQKNV 300
```


61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGGLDILNTQIGSPAFRVEVQADG 118
 60 ADSRASDIPGLPTNPLRFAA--SEVSLHGALEVLHDKGGLDILNTQIGSPAFRVEVQADG 117
 119 THAAIGKNGLEVSVTLSPOEWSLSQIDTEGKRNRFVFTGGRGGCHPMVTVASDIAEAR 178
 118 SHVAIGKNGLETTVVLSSQESLSQSLDPECKNKFVFTGGRGGAGHAMVTVASDIAEAR 177
 179 TRILAKLDP-DNHGGRQPKVDVTRSVGVSGASGIDGCVVSETHSTTNSVSRSDPKFWVS 237
 178 QRILDKLEPKDKETKEPDGPN--GEKIIETHTSTSSLRADPKLWLS 226
 238 VGAIAGLAGLAATGTAQALALTPEPDDPTTDPQAAANAESATKDQLTQEAFAKNPENO 297
 227 LGTIAAGLIGMAATGTAQAVALTPEPDDPTTDPDAANTAAAKQDLTKEAFQPNQ 286
 298 KVNIDANGNAIPSGELXDDIVEQIAQAQAEVAVQAVESNAQAQRYEDOHARROE 357
 287 KVNIDENGNAIPSGELKDDVVAQIAEQAKAAGEARQEAIESNSQAQRYEDOHAKREQE 346
 358 LQSSGIGYGLSALIVAGGIVTALHRRNQPAEQTTTTHTVVQOQTGGIPQHKV 417
 347 MSLSSGVGYGIGSALILGGIGAGVTAALHRKNQPAEQ--TITRTVVDNQ----PTNNA 400
 418 ALM-----POERRRFRDRDSQGSVASTHWSDS--EVVNPYAEVGGARNLSLAHOP 468
 401 SAQGNNTDTSPEESP--ASRRNSNASLAS-NGSDTSTSTGTVENPYADVGMPRNDLSLARIS 457
 469 EHHYDEVAADPGYVIONFSGSPVTRGLTGPQGIQSTYALLANSGLRLGMGLTGS 528
 458 EEPYDEVAADPNYSVIOHFSGNSPVTGRLVGTGPQGIQSTYALLANSGLRLGMGLTGS 517
 529 GGETAVSSVNAAPTGPVRFV 549
 518 GGESAVSTANAAPTGPAREV 538

RESULT 8
 Q47014
 ID Q47014 PRELIMINARY; PRT; 538 AA.
 AC Q47014;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE TRANSLOCATED INTIMIN RECEPTOR TIR.
 GN TIR.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RN [2]
 RN [3]
 RN [4]
 RN [5]
 RN [6]
 RN [7]
 RN [8]
 RN [9]
 RN [10]
 RN [11]
 RN [12]
 RN [13]
 RN [14]
 RN [15]
 RN [16]
 RN [17]
 RN [18]
 RN [19]
 RN [20]
 RN [21]
 RN [22]
 RN [23]
 RN [24]
 RN [25]
 RN [26]
 RN [27]
 RN [28]
 RN [29]
 RN [30]
 RN [31]
 RN [32]
 RN [33]
 RN [34]
 RN [35]
 RN [36]
 RN [37]
 RN [38]
 RN [39]
 RN [40]
 RN [41]
 RN [42]
 RN [43]
 RN [44]
 RN [45]
 RN [46]
 RN [47]
 RN [48]
 RN [49]
 RN [50]
 RN [51]
 RN [52]
 RN [53]
 RN [54]
 RN [55]
 RN [56]
 RN [57]
 RN [58]
 RN [59]
 RN [60]
 RN [61]
 RN [62]
 RN [63]
 RN [64]
 RN [65]
 RN [66]
 RN [67]
 RN [68]
 RN [69]
 RN [70]
 RN [71]
 RN [72]
 RN [73]
 RN [74]
 RN [75]
 RN [76]
 RN [77]
 RN [78]
 RN [79]
 RN [80]
 RN [81]
 RN [82]
 RN [83]
 RN [84]
 RN [85]
 RN [86]
 RN [87]
 RN [88]
 RN [89]
 RN [90]
 RN [91]
 RN [92]
 RN [93]
 RN [94]
 RN [95]
 RN [96]
 RN [97]
 RN [98]
 RN [99]
 RN [100]
 RN [101]
 RN [102]
 RN [103]
 RN [104]
 RN [105]
 RN [106]
 RN [107]
 RN [108]
 RN [109]
 RN [110]
 RN [111]
 RN [112]
 RN [113]
 RN [114]
 RN [115]
 RN [116]
 RN [117]
 RN [118]
 RN [119]
 RN [120]
 RN [121]
 RN [122]
 RN [123]
 RN [124]
 RN [125]
 RN [126]
 RN [127]
 RN [128]
 RN [129]
 RN [130]
 RN [131]
 RN [132]
 RN [133]
 RN [134]
 RN [135]
 RN [136]
 RN [137]
 RN [138]
 RN [139]
 RN [140]
 RN [141]
 RN [142]
 RN [143]
 RN [144]
 RN [145]
 RN [146]
 RN [147]
 RN [148]
 RN [149]
 RN [150]
 RN [151]
 RN [152]
 RN [153]
 RN [154]
 RN [155]
 RN [156]
 RN [157]
 RN [158]
 RN [159]
 RN [160]
 RN [161]
 RN [162]
 RN [163]
 RN [164]
 RN [165]
 RN [166]
 RN [167]
 RN [168]
 RN [169]
 RN [170]
 RN [171]
 RN [172]
 RN [173]
 RN [174]
 RN [175]
 RN [176]
 RN [177]
 RN [178]
 RN [179]
 RN [180]
 RN [181]
 RN [182]
 RN [183]
 RN [184]
 RN [185]
 RN [186]
 RN [187]
 RN [188]
 RN [189]
 RN [190]
 RN [191]
 RN [192]
 RN [193]
 RN [194]
 RN [195]
 RN [196]
 RN [197]
 RN [198]
 RN [199]
 RN [200]
 RN [201]
 RN [202]
 RN [203]
 RN [204]
 RN [205]
 RN [206]
 RN [207]
 RN [208]
 RN [209]
 RN [210]
 RN [211]
 RN [212]
 RN [213]
 RN [214]
 RN [215]
 RN [216]
 RN [217]
 RN [218]
 RN [219]
 RN [220]
 RN [221]
 RN [222]
 RN [223]
 RN [224]
 RN [225]
 RN [226]
 RN [227]
 RN [228]
 RN [229]
 RN [230]
 RN [231]
 RN [232]
 RN [233]
 RN [234]
 RN [235]
 RN [236]
 RN [237]
 RN [238]
 RN [239]
 RN [240]
 RN [241]
 RN [242]
 RN [243]
 RN [244]
 RN [245]
 RN [246]
 RN [247]
 RN [248]
 RN [249]
 RN [250]
 RN [251]
 RN [252]
 RN [253]
 RN [254]
 RN [255]
 RN [256]
 RN [257]
 RN [258]
 RN [259]
 RN [260]
 RN [261]
 RN [262]
 RN [263]
 RN [264]
 RN [265]
 RN [266]
 RN [267]
 RN [268]
 RN [269]
 RN [270]
 RN [271]
 RN [272]
 RN [273]
 RN [274]
 RN [275]
 RN [276]
 RN [277]
 RN [278]
 RN [279]
 RN [280]
 RN [281]
 RN [282]
 RN [283]
 RN [284]
 RN [285]
 RN [286]
 RN [287]
 RN [288]
 RN [289]
 RN [290]
 RN [291]
 RN [292]
 RN [293]
 RN [294]
 RN [295]
 RN [296]
 RN [297]
 RN [298]
 RN [299]
 RN [300]
 RN [301]
 RN [302]
 RN [303]
 RN [304]
 RN [305]
 RN [306]
 RN [307]
 RN [308]
 RN [309]
 RN [310]
 RN [311]
 RN [312]
 RN [313]
 RN [314]
 RN [315]
 RN [316]
 RN [317]
 RN [318]
 RN [319]
 RN [320]
 RN [321]
 RN [322]
 RN [323]
 RN [324]
 RN [325]
 RN [326]
 RN [327]
 RN [328]
 RN [329]
 RN [330]
 RN [331]
 RN [332]
 RN [333]
 RN [334]
 RN [335]
 RN [336]
 RN [337]
 RN [338]
 RN [339]
 RN [340]
 RN [341]
 RN [342]
 RN [343]
 RN [344]
 RN [345]
 RN [346]
 RN [347]
 RN [348]
 RN [349]
 RN [350]
 RN [351]
 RN [352]
 RN [353]
 RN [354]
 RN [355]
 RN [356]
 RN [357]
 RN [358]
 RN [359]
 RN [360]
 RN [361]
 RN [362]
 RN [363]
 RN [364]
 RN [365]
 RN [366]
 RN [367]
 RN [368]
 RN [369]
 RN [370]
 RN [371]
 RN [372]
 RN [373]
 RN [374]
 RN [375]
 RN [376]
 RN [377]
 RN [378]
 RN [379]
 RN [380]
 RN [381]
 RN [382]
 RN [383]
 RN [384]
 RN [385]
 RN [386]
 RN [387]
 RN [388]
 RN [389]
 RN [390]
 RN [391]
 RN [392]
 RN [393]
 RN [394]
 RN [395]
 RN [396]
 RN [397]
 RN [398]
 RN [399]
 RN [400]
 RN [401]
 RN [402]
 RN [403]
 RN [404]
 RN [405]
 RN [406]
 RN [407]
 RN [408]
 RN [409]
 RN [410]
 RN [411]
 RN [412]
 RN [413]
 RN [414]
 RN [415]
 RN [416]
 RN [417]
 RN [418]
 RN [419]
 RN [420]
 RN [421]
 RN [422]
 RN [423]
 RN [424]
 RN [425]
 RN [426]
 RN [427]
 RN [428]
 RN [429]
 RN [430]
 RN [431]
 RN [432]
 RN [433]
 RN [434]
 RN [435]
 RN [436]
 RN [437]
 RN [438]
 RN [439]
 RN [440]
 RN [441]
 RN [442]
 RN [443]
 RN [444]
 RN [445]
 RN [446]
 RN [447]
 RN [448]
 RN [449]
 RN [450]
 RN [451]
 RN [452]
 RN [453]
 RN [454]
 RN [455]
 RN [456]
 RN [457]
 RN [458]
 RN [459]
 RN [460]
 RN [461]
 RN [462]
 RN [463]
 RN [464]
 RN [465]
 RN [466]
 RN [467]
 RN [468]
 RN [469]
 RN [470]
 RN [471]
 RN [472]
 RN [473]
 RN [474]
 RN [475]
 RN [476]
 RN [477]
 RN [478]
 RN [479]
 RN [480]
 RN [481]
 RN [482]
 RN [483]
 RN [484]
 RN [485]
 RN [486]
 RN [487]
 RN [488]
 RN [489]
 RN [490]
 RN [491]
 RN [492]
 RN [493]
 RN [494]
 RN [495]
 RN [496]
 RN [497]
 RN [498]
 RN [499]
 RN [500]
 RN [501]
 RN [502]
 RN [503]
 RN [504]
 RN [505]
 RN [506]
 RN [507]
 RN [508]
 RN [509]
 RN [510]
 RN [511]
 RN [512]
 RN [513]
 RN [514]
 RN [515]
 RN [516]
 RN [517]
 RN [518]
 RN [519]
 RN [520]
 RN [521]
 RN [522]
 RN [523]
 RN [524]
 RN [525]
 RN [526]
 RN [527]
 RN [528]
 RN [529]
 RN [530]
 RN [531]
 RN [532]
 RN [533]
 RN [534]
 RN [535]
 RN [536]
 RN [537]
 RN [538]
 RN [539]
 RN [540]
 RN [541]
 RN [542]
 RN [543]
 RN [544]
 RN [545]
 RN [546]
 RN [547]
 RN [548]
 RN [549]
 RN [550]
 RN [551]
 RN [552]
 RN [553]
 RN [554]
 RN [555]
 RN [556]
 RN [557]
 RN [558]
 RN [559]
 RN [560]
 RN [561]
 RN [562]
 RN [563]
 RN [564]
 RN [565]
 RN [566]
 RN [567]
 RN [568]
 RN [569]
 RN [570]
 RN [571]
 RN [572]
 RN [573]
 RN [574]
 RN [575]
 RN [576]
 RN [577]
 RN [578]
 RN [579]
 RN [580]
 RN [581]
 RN [582]
 RN [583]
 RN [584]
 RN [585]
 RN [586]
 RN [587]
 RN [588]
 RN [589]
 RN [590]
 RN [591]
 RN [592]
 RN [593]
 RN [594]
 RN [595]
 RN [596]
 RN [597]
 RN [598]
 RN [599]
 RN [600]
 RN [601]
 RN [602]
 RN [603]
 RN [604]
 RN [605]
 RN [606]
 RN [607]
 RN [608]
 RN [609]
 RN [610]
 RN [611]
 RN [612]
 RN [613]
 RN [614]
 RN [615]
 RN [616]
 RN [617]
 RN [618]
 RN [619]
 RN [620]
 RN [621]
 RN [622]
 RN [623]
 RN [624]
 RN [625]
 RN [626]
 RN [627]
 RN [628]
 RN [629]
 RN [630]
 RN [631]
 RN [632]
 RN [633]
 RN [634]
 RN [635]
 RN [636]
 RN [637]
 RN [638]
 RN [639]
 RN [640]
 RN [641]
 RN [642]
 RN [643]
 RN [644]
 RN [645]
 RN [646]
 RN [647]
 RN [648]
 RN [649]
 RN [650]
 RN [651]
 RN [652]
 RN [653]
 RN [654]
 RN [655]
 RN [656]
 RN [657]
 RN [658]
 RN [659]
 RN [660]
 RN [661]
 RN [662]
 RN [663]
 RN [664]
 RN [665]
 RN [666]
 RN [667]
 RN [668]
 RN [669]
 RN [670]
 RN [671]
 RN [672]
 RN [673]
 RN [674]
 RN [675]
 RN [676]
 RN [677]
 RN [678]
 RN [679]
 RN [680]
 RN [681]
 RN [682]
 RN [683]
 RN [684]
 RN [685]
 RN [686]
 RN [687]
 RN [688]
 RN [689]
 RN [690]
 RN [691]
 RN [692]
 RN [693]
 RN [694]
 RN [695]
 RN [696]
 RN [697]
 RN [698]
 RN [699]
 RN [700]
 RN [701]
 RN [702]
 RN [703]
 RN [704]
 RN [705]
 RN [706]
 RN [707]
 RN [708]
 RN [709]
 RN [710]
 RN [711]
 RN [712]
 RN [713]
 RN [714]
 RN [715]
 RN [716]
 RN [717]
 RN [718]
 RN [719]
 RN [720]
 RN [721]
 RN [722]
 RN [723]
 RN [724]
 RN [725]
 RN [726]
 RN [727]
 RN [728]
 RN [729]
 RN [730]
 RN [731]
 RN [732]
 RN [733]
 RN [734]
 RN [735]
 RN [736]
 RN [737]
 RN [738]
 RN [739]
 RN [740]
 RN [741]
 RN [742]
 RN [743]
 RN [744]
 RN [745]
 RN [746]
 RN [747]
 RN [748]
 RN [749]
 RN [750]
 RN [751]
 RN [752]
 RN [753]
 RN [754]
 RN [755]
 RN [756]
 RN [757]
 RN [758]
 RN [759]
 RN [760]
 RN [761]
 RN [762]
 RN [763]
 RN [764]
 RN [765]
 RN [766]
 RN [767]
 RN [768]
 RN [769]
 RN [770]
 RN [771]
 RN [772]
 RN [773]
 RN [774]
 RN [775]
 RN [776]
 RN [777]
 RN [778]
 RN [779]
 RN [780]
 RN [781]
 RN [782]
 RN [783]
 RN [784]
 RN [785]
 RN [786]
 RN [787]
 RN [788]
 RN [789]
 RN [790]
 RN [791]
 RN [792]
 RN [793]
 RN [794]
 RN [795]
 RN [796]
 RN [797]
 RN [798]
 RN [799]
 RN [800]
 RN [801]
 RN [802]
 RN [803]
 RN [804]
 RN [805]
 RN [806]
 RN [807]
 RN [808]
 RN [809]
 RN [810]
 RN [811]
 RN [812]
 RN [813]
 RN [814]
 RN [815]
 RN [816]
 RN [817]
 RN [818]
 RN [819]
 RN [820]
 RN [821]
 RN [822]
 RN [823]
 RN [824]
 RN [825]
 RN [826]
 RN [827]
 RN [828]
 RN [829]
 RN [830]
 RN [831]
 RN [832]
 RN [833]
 RN [834]
 RN [835]
 RN [836]
 RN [837]
 RN [838]
 RN [839]
 RN [840]
 RN [841]
 RN [842]
 RN [843]
 RN [844]
 RN [845]
 RN [846]
 RN [847]
 RN [848]
 RN [849]
 RN [850]
 RN [851]
 RN [852]
 RN [853]
 RN [854]
 RN [855]
 RN [856]
 RN [857]
 RN [858]
 RN [859]
 RN [860]
 RN [861]
 RN [862]
 RN [863]
 RN [864]
 RN [865]
 RN [866]
 RN [867]
 RN [868]
 RN [869]
 RN [870]
 RN [871]
 RN [872]
 RN [873]
 RN [874]
 RN [875]
 RN [876]
 RN [877]
 RN [878]
 RN [879]
 RN [880]
 RN [881]
 RN [882]
 RN [883]
 RN [884]
 RN [885]
 RN [886]
 RN [887]
 RN [888]
 RN [889]
 RN [890]
 RN [891]
 RN [892]
 RN [893]
 RN [894]
 RN [895]
 RN [896]
 RN [897]
 RN [898]
 RN [899]
 RN [900]
 RN [901]
 RN [902]
 RN [903]
 RN [904]
 RN [905]
 RN [906]
 RN [907]
 RN [908]
 RN [909]
 RN [910]
 RN [911]
 RN [912]
 RN [913]
 RN [914]
 RN [915]
 RN [916]
 RN [917]
 RN [918]
 RN [919]
 RN [920]
 RN [921]
 RN [922]
 RN [923]
 RN [924]
 RN [925]
 RN [926]
 RN [927]
 RN [928]
 RN [929]
 RN [930]
 RN [931]
 RN [932]
 RN [933]
 RN [934]
 RN [935]
 RN [936]
 RN [937]
 RN [938]
 RN [939]
 RN [940]
 RN [941]
 RN [942]
 RN [943]
 RN [944]
 RN [945]
 RN [946]
 RN [947]
 RN [948]
 RN [949]
 RN [950]
 RN [951]
 RN [952]
 RN [953]
 RN [954]
 RN [955]
 RN [956]
 RN [957]
 RN [958]
 RN [959]
 RN [960]
 RN [961]
 RN [962]
 RN [963]
 RN [964]
 RN [965]
 RN [966]
 RN [967]
 RN [968]
 RN [969]
 RN [970]
 RN [971]
 RN [972]
 RN [973]
 RN [974]
 RN [975]
 RN [976]
 RN [977]
 RN [978]
 RN [979]
 RN [980]
 RN [981]
 RN [982]
 RN [983]
 RN [984]
 RN [985]
 RN [986]
 RN [987]
 RN [988]
 RN [989]
 RN [990]
 RN [991]
 RN [992]
 RN [993]
 RN [994]
 RN [995]
 RN [996]
 RN [997]
 RN [998]
 RN [999]
 RN [1000]

sequence heterogeneity.*;
 RL Infect. Immun. 66:5580-5586(1998).
 RP [4]
 RC STRAIN-B10;
 RA Marches O., Nougayrede J.-P., Boury M., Mainil J., Charlier G.,
 RA Boullier S., De Rycke J., Milon A., Oswald E.;
 RA "Role of tir and intimin in the pathogenesis of rabbit
 enteropathogenic Escherichia coli.";
 RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RL EMBL; U59502; AAC32028.2;
 DR EMBL; AF070068; AAC1065.1;
 DR EMBL; AF070068; AAC69316.1;
 DR EMBL; AF132728; AAD27868.1;
 DR EMBL; AF113597; AAF03080.1;
 DR InterPro; IPR003536;
 DR PRINTS; PRO1370; TRNSINTIMINR.
 KW Hypothetical protein; Receptor.
 SQ SEQUENCE 538 AA; 55420 MW; 31D7A8E227B3D06C CRC64;
 Query Match 63.6%; Score 1781.5; DB 2; Length 538;
 Best Local Similarity 65.4%; Pred. No. 5.1e-96;
 Matches 367; Conservative 66; Mismatches 93; Indels 35; Gaps 11;
 QY 1 MTGILGNVNGNHLIPPPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
 DB 1 MTGILGNVNGNHLIPPPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 59
 QY 61 VDSR--DIPGLPTNPSRLAAATSETCLLGGFEVLHDKGGLDILNTQIGSPAFRVEVQADG 118
 DB 60 ADSRASDIPGLPTNPLRFAA--SEVSLHGALEVLHDKGGLDILNTQIGSPAFRVEVQADG 117
 QY 119 THAAIGKNGLEVSVTLSPOEWSLSQIDTEGKRNRFVFTGGRGGCHPMVTVASDIAEAR 178
 DB 118 SHVAIGKNGLETTVVLSSQESLSQSLDPECKNKFVFTGGRGGAGHAMVTVASDIAEAR 177
 QY 179 TRILAKLDP-DNHGGRQPKVDVTRSVGVSGASGIDGCVVSETHSTTNSVSRSDPKFWVS 237
 DB 178 QRILDKLEPKDKETKEPDGPN--GEKIIETHTSTSSLRADPKLWLS 226
 QY 238 VGAIAGLAGLAATGTAQALALTPEPDDPTTDPQAAANAESATKDQLTQEAFAKNPENO 297
 DB 227 LGTIAAGLIGMAATGTAQAVALTPEPDDPTTDPDAANTAAAKQDLTKEAFQPNQ 286
 QY 298 KVNIDANGNAIPSGELXDDIVEQIAQAQAEVAVQAVESNAQAQRYEDOHARROE 357
 DB 287 KVNIDENGNAIPSGELKDDVVAQIAEQAKAAGEARQEAIESNSQAQRYEDOHAKREQE 346
 QY 358 LQSSGIGYGLSALIVAGGIVTALHRRNQPAEQTTTTHTVVQOQTGGIPQHKV 417
 DB 347 MSLSSGVGYGIGSALILGGIGAGVTAALHRKNQPAEQ--TITRTVVDNQ----PTNNA 400
 QY 418 ALM-----POERRRFRDRDSQGSVASTHWSDS--EVVNPYAEVGGARNLSLAHOP 468
 DB 401 SAQGNNTDTSPEESP--ASRRNSNASLAS-NGSDTSTSTGTVENPYADVGMPRNDLSLARIS 457
 QY 469 EHHYDEVAADPGYVIONFSGSPVTRGLTGPQGIQSTYALLANSGLRLGMGLTGS 528
 DB 458 EEPYDEVAADPNYSVIOHFSGNSPVTGRLVGTGPQGIQSTYALLANSGLRLGMGLTGS 5

| | | | |
|--------|---|---------------------------|---|
| DE | TRANSLOCATED INTIMIN RECEPTOR TIR. | DE | TRANSLOCATED INTIMIN RECEPTOR TIR. |
| GN | TIR. | GN | TIR. |
| OS | Escherichia coli. | OS | Escherichia coli. |
| OC | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; | OC | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; |
| OC | Escherichia. | OC | Escherichia. |
| OX | NCBI_TaxID=562; | OX | NCBI_TaxID=562; |
| RP | (1) | RP | (1) |
| RP | SEQUENCE FROM N.A. | RP | SEQUENCE FROM N.A. |
| RC | STRAIN=EP87A; | RC | STRAIN=86/24; |
| RA | Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.: | RA | MEDLINE=99242825; PubMed=10225900; |
| RT | "The translocated intimin receptors (Tir) of Shiga toxinigenic | RA | Devlinney R., Stein M., Reinscheid D., Abe A., Ruschkowski S., |
| RT | Escherichia coli isolates belonging to serogroups O26, O111, and O157 | RA | Finlay B.B.; |
| RT | react with sera from patients with hemolytic-uremic syndrome and | RT | "Enterohemorrhagic Escherichia coli O157:H7 produces Tir, which is |
| RT | exhibit marked sequence heterogeneity." | RT | translocated to the host cell membrane but is not tyrosine |
| RL | Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases. | RL | phosphorylated." |
| DR | EMBL; AF070069; AAC69318.1; - | RL | Infect. Immun. 67:2389-2398(1999). |
| DR | InterPro; IPR003536; - | RN | (2) |
| DR | PRINTS; PR01370; TRNSINTMINR. | RP | SEQUENCE FROM N.A. |
| SQ | SEQUENCE 538 AA; 55602 MW; 447052A0E3214D6D CRC64; | RC | STRAIN=ED1933; |
| | | RC | MEDLINE=98339885; PubMed=9673266; |
| | | RA | Perna N.T., Mayhew G.F., Posfal G., Elliott S., Donnenberg M.S., |
| | | RA | Kaper J.B., Blattner F.R.; |
| | | RT | "Molecular evolution of a pathogenicity island from enterohemorrhagic |
| | | RT | Escherichia coli O157:H7." |
| | | RL | Infect. Immun. 66:3810-3817(1998). |
| | | DR | EMBL; AF125993; AAD29391.1; - |
| | | DR | EMBL; AF071034; AAC31506.1; - |
| | | DR | InterPro; IPR003536; - |
| | | DR | PRINTS; PR01370; TRNSINTMINR. |
| | | KW | Receptor. |
| | | SQ | SEQUENCE 558 AA; 58022 MW; 99C417222D4B4AA1 CRC64; |
| | | | |
| | | Query Match | 54.6%; Score 1528.5; DB 2; Length 558; |
| | | Best Local Similarity | 56.2%; Pred. No 2.7e-81; |
| | | Matches 327; Conservative | 60; Mismatches 138; Indels 57; Gaps 11; |
| QY | 1 MPICNLGNVNGNHLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPRLNSMADS 60 | | |
| Db | 1 MPICNLGNHNPVNRALIPAPPLPSQTDGAG-GARNQLINSNGPMGSRLLTPIRNSVADA 59 | | |
| QY | 61 VDNR--DIPGLTPNPSRLAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEVQADG 118 | | |
| Db | 60 ADSRASDIPGLTPNPLFAA--SEVSLHGAEVLHDKGGLDTLNSAIGSLFVRVETRDDG 117 | | |
| QY | 119 THAATGKNGLEVSVTLSPQWSSLSQSDTEGKNRFVFTGGRGSGHPMTVVASDIAEAR 178 | | |
| Db | 118 SHVALGQKNGLETIVVLSQDEFFSIQSLDPEGKNKFVFTGGRGSGAGHAMTVASDIAEAR 177 | | |
| QY | 179 TRIAKLDP--DNHGGKQPKVDYTRSVGVSASGIDDDGVVSTHTSTNNSVRSRDPKFWVS 237 | | |
| Db | 178 QRIIDKLEPKDTKETKEGDPNS-----GEGKIIIEHTSTSTFLSRADPKLWLS 226 | | |
| QY | 238 VCAIAGLAGLAATGIAQALATPEPDDPTTDPDOAANAASATKDOLTQEAFAKNPENQ 297 | | |
| Db | 227 LGTIAAGLIGMAATGIAQALATPEPDDPTTDPDTAASTAAATKRLTQEAQDPDKQ 286 | | |
| QY | 298 KVNIDANGNALPSGELXDDIVEQIAQAQAKEAGEVARQAQAVESNAQAQORYEDQHARQEE 357 | | |
| Db | 287 KVNIDENGNALPSGELIDVVVAQIAEQAKAAGEAQAEIESNSQAQKKYDEQAKRQE 346 | | |
| QY | 358 LQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTHTVVQQQTGGIPQHKV 417 | | |
| Db | 347 MSLSGCVGYSIGALILGGIGAGVTAALHRRNQPAEQ--TITTVVDNQ-----PTNNA 400 | | |
| QY | 418 ALM-----PQERRFRDRDSQGSVASTHWSDS--EVNPNYAEGVGNARNSLSAHQP 468 | | |
| Db | 401 SAQNGNTDTSGPESP--ASRRNSNASLAS--NGSDTSSTGTVENPADVGMPRNDSLARI 457 | | |
| QY | 469 EEHIVDEVAADPGYSVIQNFSGSPVTRIGTIPGQGIQSTYALLANGSGLRGMGGLTS 528 | | |
| Db | 458 EEPIYDEVAADPNYSVIQHFSGNSPVTGRVLGTPGQGIQSTYALLASSGGLRGMGGLTG 517 | | |
| QY | 529 GGETAVSSVNAAPTGPVRFV 549 | | |
| Db | 518 GGSAYSTANASPTGPARFV 538 | | |
| RESULT | 10 | | |
| Q9R396 | PRELIMINARY; PRT: 558 AA. | | |
| ID | Q9R396 | | |
| AC | Q9R396; | | |
| DT | 01-MAY-2000 (TReMBLrel. 13, Created) | | |
| DT | 01-MAY-2000 (TReMBLrel. 13, Last sequence update) | | |
| DT | 01-MAR-2001 (TReMBLrel. 16, Last annotation update) | | |

Db 540 TADRISSNGGGAIGSL--SHSLAHKVPSPSSAAAASRLVEYHHHHQHVSPKRILREF 597
QY 474 DEVAADPGYVIONFNGSGPVTGRLICTPGQGIQSTYALLANSGLRGMGLTS--GGE 531
Db 598 EKVSLDNNVCNNGSGV-----GGISGGAGGKRRAKGS 633
QY 532 TAVSSV-----NAAP-----TPGP 545
Db 634 TATSAVTTKSPINLAPPQAKVLSTPTT 661

RESULT 13
Q9N1P0 ID Q9N1P0 PRELIMINARY; PRT; 818 AA.
AC Q9N1P0;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE SUBMAXILIARY MUCIN (FRAGMENT).
GN BSM1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BREED ANGUS;
RX MEDLINE=2023253; PubMed=10759843;
RA Jang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.P.;
RT "The central domain of bovine submaxillary mucin consists of over 50
RT tandem repeats of 329 amino acids: chromosomal localization of the
RT BSM1 gene and relations to ovine and porcine counterparts";
RL Eur. J. Biochem. 267:2208-2217(2000).
DR EMBL: AF178428; AAF67279.1;
FT NON_TER 818
FT NON_TER 818
SQ SEQUENCE 818 AA; 73014 MW; B5BB44F84F66F86B CRC64;

Query Match 5.7%; Score 160.5; DB 6; Length 818;
Best Local Similarity 21.68; Pred. No. 0.15;
Matches 122; Conservative 64; Mismatches 221; Indels 157; Gaps 21;

QY 10 VGNHLPAPPPLPSQTDGAARGTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGL 69
Db 174 VSGAAVPGSPGRSRATAVSGEGSQPTVALSGATGTSAGPSGRST-----SSAIPAT 227
QY 70 PTNPSRLAAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEQADGTHAATGEKNGL 129
Db 228 PGSTTGAAGAGTGA-----VDSQQTASLPAAARPTALGPSTAPSGTSES 274
QY 130 EVSVTLSPQWSSLSQSDTEGKRFVFTGGRGGHGMVTVASDIAEARTI---LAKL 185
Db 275 RSVPP-----GGSETTQPG-----AGSEPTLSPGVTTTALRGSETRVPSTGVSG 322
QY 186 DPNHGGQRPQKVDTRSVGSGASIDGVVSETHST---TNSSVSRDPPKFWVSVGAIA 242
Db 323 PGSTGG-----SAATGSGAGSGPTAPVSGETRTSVISGNTVPVSGAP---VTPGSSA 373
QY 243 -----AGLAGLATAIAQALALTEPPDDPTTDDQANRAESATK 283
Db 374 GSSGAPGTGGPGSETASPLSGAAGTSATGSGTSI---PPSGAPVTPPEPLISTGA-SAGP 429
QY 284 DQLTQEAFFKPNKVNNDANGNATP-SGELKDDIVEQIAQAKEAGEVRAQQAESNAQ 342
Db 430 PASSESTVTLPGATGTDVLRSGTSLPVS-----GAVTPAPSPGGSSA----- 472
QY 343 AQORYEDQHARRQEBQLSSGIGYGLSSALIVAGGIGAGVTALHRRNQPAEQTTTTTH 402
Db 473 -----TAGPGVGSATTVQASGATGADV----- 495
QY 403 TVVQOQTGCIPOHKVAMP--QERRRDRDRDSQGSVASTHSDSSSEVNPVYAEVGGAR 460

Db 496 ----RSTSLPVGVAVPCSGSPCRSGATAVSSQSGPTVALSGATGTSVP-----SGTR 547
QY 461 NLSAQHQPHEHIYDEVAADPGYVIONFNGSGPVTGRLICTPGQGIQSTYALLANSGLR 520
Db 548 FSSSA-----IPATPG-----STTGRAAG--AGTPGVDSQQTASLPAAARPTA 588
QY 521 LGMGGLTSGGETAVSSVNAAPTGP 544
Db 589 LGPQTSAPSGET---SESRSVPG 609

RESULT 14
Q9FDA0 ID Q9FDA0 PRELIMINARY; PRT; 1265 AA.
AC Q9FDA0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN XADA.
GN XADA.
OS Xanthomonas oryzae pv. oryzae.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=64187;
RN [1]
RP SEQUENCE FROM N.A.
RA Ray S.K., Rajeshwari R., Sonti R.V.;
RT "A putative outer membrane protein from Xanthomonas oryzae pv. oryzae
RT that is involved in virulence";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF288222; AAG01335.1;
SQ SEQUENCE 1265 AA; 119856 MW; 21762579B5EC70A2 CRC64;

Query Match 5.7%; Score 159.5; DB 2; Length 1265;
Best Local Similarity 23.48; Pred. No. 0.3;
Matches 136; Conservative 73; Mismatches 229; Indels 143; Gaps 29;

QY 3 IGNLGNVNGNHLIPAPPPLPSQTDGAARGTGHLISS-TGALGSRSLFSPLR-----N 55
Db 592 LGNTARALGDNSSAVGSNAVADIGATANGAGAQAALSAYTIALGSKAVASDLQAIEGN 651
QY 56 SMADSVDSRDIPGLPTNPRLAAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEQ 115
Db 652 STAGNVGSTALGGFSQSGRLUSSALGYSAVSSVD-----STAVG-----VAAQ 695
QY 116 ADGTHA-AIGE--KNGLEVSVTLSPQWSSLSQSDTEGKRFVFTGGRGGHGMVTVAS 172
Db 696 ATGVSVAIGAELSKATGESVAVGGAFSGWPTQASCKGAAAFAGAWATADYTTAIGR 755
QY 173 D-IAEA--RTRILAKLD--PDN-----HGGROPKQVDTRSVGV-GSASGIDD-----GV 215
Db 756 DSYADGVNATAVAGQSADALADNTLALGGSRRAKAVGASVIGVDASATGINSTGVGRQVNV 815
QY 216 VSEHTST-TNSSVSRDPKFWVSVCAIAAGLAGLATAIAQALALTEPPDDPTTDDQA 274
Db 816 IGENAVSVGYNVRQS-----AVNGVALG-ANAGATG-ADSVL---GSGSSTTDADTV 865
QY 275 ANAESATKDLQTOEAFKNPENQKVN-----DANGNAIPSGELXDDIVEQIAQAKEA 328
Db 866 SVSGNGRGGPATRI-----VNVGAGAVASASTDAINGGOLFESL----- 906
QY 329 GEVARQQAQVESAQAQRYEDQHARRQEBQLSSGIGYGLSSALIVAGGIGAGVTALH- 387
Db 907 -----SNA-----ASPLGGG---AALGAGQGVFAPTYLIQG 934
QY 388 ---RRNQPAEQTTTTTHTVVQOQTGCIPOHKVAMPQERRRDRDRDSQGSVASTHSDS 445
Db 935 ASYNNVGAALTALDSKVITELDARSGGTPANTAA-----RTASLRATATVPMAAAVAVSAV 988
QY 446 SSEV-----VNPVYAEVGGARNLSAQHQPHEHIYDEVAADPGYVIONFNGSGPVTGRLICT 501

Db 989 SSVASTAIDATAGVGT-----PTAAVVGSTTAAVTVGTAANNVTGTAIG- 1039

QY 502 PCQGTQSTYALLANGGL-----RLCMGGLTS-GGTAVSSV 537

Db 1040 -----GSAYAHGANDTAIGSNARVNADGSTAVGANTQIAAV 1075

RESULT 15

Q9NHX4 Q9NHX4 PRELIMINARY; PRT: 1323 AA.

AC Q9NHX4;

DT 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)

DE ALHAMBRA.

GN ALHAMBRA OR CGL070.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_taxid=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA Perrin L., Dura J.M.;

RT "Alhambra, a Drosophila homolog of mammalian AF-10 and AF-17.";

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF217960; AAF2595.1; -.

DR FlyBase; FBgn0037471; Alhambra.

DR InterPro; IPR001965; -.

DR SMART; SM00249; PHD; 1.

SQ SEQUENCE 1323 AA; 132883 MW; D53C0C8AF392F9A6 CRC64;

Query Match 5.6%; Score 156.5; DB 5; Length 1323;

Best Local Similarity 20.2%; Pred. NO. 0.47;

Matches 135; Conservative 85; Mismatches 270; Indels 177; Gaps 25;

QY 3 ICNLGNVNGNHLIPAPLPSPQTGAARGGTGHLIS-----STGALGSRSLFSP-LRNS 56

Db 322 ICNISLNN-----LPGSSSTSSAGNVPGGGIIISASSGGGATQSTSSQSTAPGTTKS 378

QY 57 MADSVDSRDIPGLTPNPSRLAAATSETCLLGGFEVLHDKGLDILNTQIGPSAFRVEQOA 116

Db 379 SASSSSSSNSYKEXHKSLSKSTSSKDKG-----KDSSTNSANNFTNSSASSTSSNS 433

QY 117 DQTHAIGKNGLEVS-----VTLSPQEWSSLOSIDPE-GKNRFVFTGGGGG-HPM 167

Db 434 SSTR-----EKSSKLSKNKDSNQVPSATSSLSSTTSINTQPSSTSTATAGSGGTGTHVS 489

QY 168 VTVASDIAEARTRIILAKLPDNDHGGROPKVDVTRSVGVCSASGIDDGV----- 215

Db 490 SAAAGLNSAPS-----TTNEHNSNHAHTLTNTGTGAGSAAGKLOSvNSLNSSSGGFGS 542

QY 216 ----VSETHTSTNNSV-----RSDPKFWVSVGAIAAGLAGLAA--TGIAQALALTP 262

Db 543 DLRSVSTSSSVTNDSTGGFGSNSNSERENLSCAGSSASNMFGTAPGTGGVSSSAAT-- 600

QY 263 PDPPTTDPQAANAESAATKDLQTAEPKNPENKVNID---ANGNAIPSGELXDDIVE 319

Db 601 --NLSTNKGSSSSSTANSUTSTSSGSSSSSSKKRKAADSAKSTSTISGSALEDNNS 658

QY 320 QIAQ-QAKEAGEVARQAVESNAQAQRYEDQHARRQEL-----QLSSGIGYGLSSA 371

Db 659 LISRYDIKDV-----QVALTPLTDFEKEIEKSKRQTELSPPHTQTSATAEVNAPLASS 713

QY 372 LIVAGGIGAGVTTALHRRNQPABEQTT-----TTHTVWQQTG 410

Db 714 TSASTAVTASATAA---SAPPATGTTTLGSSISGNAGTSSGNSGVSATGGAQAVGSG 770

QY 411 GIPOHKVALMPQERRRFRDROSGSVASTHWS-----SSSEVWNPYAE 455

Db 771 GYPKTESSKSSGTASAGSGSGSSNTSSPKHGSNIKDISSSNQQAQASTASSSAPSLVYS 830

QY 456 V-----GGARNSLSAHOPEEHYDEVAADP----- 480

Db 831 VPLSTANVPGINLPSTSSSTSSSHSASSSSGAQSOHQOQLSNALVGPMPGGSAGAF 890

QY 481 -----GYSVIONFSG-SGPVTGRLIGTPGOGIOSTYALLANSGLRLGMGG--LTSF 529

Db 891 HGGTTSAGSSSVIQQQSGKSSPALGTLVS-----GNSGGSIISASGFPLPSG 937

QY 530 GETAVSS 536

Db 938 NLTAATTT 944

Search completed: September 27, 2001, 14:27:03

Job time: 314 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:22:24 ; Search time 31.79 Seconds
(without alignments)
1066.021 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MFIGNLGHNPVNNSTPPAP.....SNSAVNTSNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SIDSL1/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|---------------------|
| 1 | 2851 | 100.0 | 559 | 20 AAY06221 | EHEC E. coli trans |
| 2 | 1498 | 52.5 | 549 | 20 AAY06220 | EPEC E. coli trans |
| 3 | 406 | 14.2 | 107 | 21 AAB20576 | Intimin C-terminal |
| 4 | 148.5 | 5.2 | 596 | 21 AAY99408 | Human PRO1342 (UNQ) |
| 5 | 148.5 | 5.2 | 596 | 22 AAB87575 | Human PRO1342. Ho |
| 6 | 148.5 | 5.2 | 596 | 22 AAB66157 | Protein of the inv |
| 7 | 141.5 | 5.0 | 1959 | 12 AAR10562 | Mutant protease (d |
| 8 | 141.5 | 5.0 | 1962 | 12 AAR10560 | Mutant protease (K |
| 9 | 141.5 | 5.0 | 1962 | 12 AAR10561 | Mutant protease (N |
| 10 | 141.5 | 5.0 | 1962 | 12 AAR10557 | Mutant protease (A |
| 11 | 141.5 | 5.0 | 1962 | 12 AAR10558 | Mutant protease (A |

| | | | | | |
|----|-------|-----|------|-------------|--------------------|
| 12 | 141.5 | 5.0 | 1962 | 12 AAR10559 | Mutant protease (A |
| 13 | 141.5 | 5.0 | 1962 | 12 AAR10563 | Mutant protease (K |
| 14 | 141.5 | 5.0 | 1968 | 12 AAR10941 | Mutant protease (d |
| 15 | 141.5 | 5.0 | 1974 | 12 AAR10940 | Mutant protease (d |
| 16 | 140.5 | 4.9 | 1026 | 21 AAY83025 | Staufen protein of |
| 17 | 134.5 | 4.7 | 1638 | 20 AAY00138 | Enterococcus faeca |
| 18 | 134.5 | 4.7 | 1638 | 20 AAY00140 | Enterococcus faeca |
| 19 | 134.5 | 4.7 | 1638 | 20 AAY00142 | Enterococcus faeca |
| 20 | 134 | 4.7 | 2514 | 21 AAY75097 | Neisseria meningit |
| 21 | 133.5 | 4.7 | 2870 | 21 AAY95559 | Caenorhabditis ele |
| 22 | 133.5 | 4.7 | 3178 | 21 AAY95556 | Caenorhabditis ele |
| 23 | 133 | 4.7 | 288 | 20 AAY29081 | T. gondii immunoge |
| 24 | 133 | 4.7 | 288 | 20 AAY29082 | T. gondii immunoge |
| 25 | 133 | 4.7 | 2314 | 22 AAB69136 | M. catarrhalis les |
| 26 | 132.5 | 4.6 | 461 | 12 AAR14530 | Usp45 protein. La |
| 27 | 132.5 | 4.6 | 550 | 20 AAY35195 | Amino acid sequenc |
| 28 | 129 | 4.5 | 1048 | 18 AAW27277 | Human cytomegalovi |
| 29 | 127.5 | 4.5 | 558 | 20 AAY00143 | Enterococcus faeca |
| 30 | 125.5 | 4.4 | 461 | 12 AAR14150 | MSP encoded by pUC |
| 31 | 124.5 | 4.4 | 1529 | 14 AAR41732 | High molecular wei |
| 32 | 124.5 | 4.4 | 2048 | 21 AAY75096 | Neisseria gonorrhe |
| 33 | 124 | 4.3 | 359 | 19 AAW81703 | M. tuberculosis im |
| 34 | 124 | 4.3 | 359 | 19 AAW64336 | Mycobacterium tube |
| 35 | 124 | 4.3 | 359 | 20 AAY39133 | M. tuberculosis an |
| 36 | 124 | 4.3 | 359 | 20 AAY38990 | M. tuberculosis re |
| 37 | 123.5 | 4.3 | 1601 | 18 AAW30292 | Non-typeable Haemo |
| 38 | 123 | 4.3 | 30 | 20 AAY06213 | EPEC E. coli trans |
| 39 | 123 | 4.3 | 1098 | 17 AAR99392 | Haemophilus adhesi |
| 40 | 123 | 4.3 | 2035 | 15 AAR57141 | Host cell factor p |
| 41 | 122.5 | 4.3 | 496 | 20 AAY37175 | Amino acid sequenc |
| 42 | 121.5 | 4.3 | 456 | 21 AAY67238 | Amino acid sequenc |
| 43 | 121.5 | 4.3 | 1992 | 17 AAW04505 | Moraxella 200 kDa |
| 44 | 121.5 | 4.3 | 1992 | 22 AAB69133 | M. catarrhalis str |
| 45 | 121.5 | 4.3 | 1992 | 22 AAB69137 | M. catarrhalis M56 |

ALIGNMENTS

RESULT 1
AAY06221
ID AAY06221 standard; Protein; 559 AA.
XX AC AAY06221;
XX AC AAY06221;
DT 16-AUG-1999 (first entry)
XX EHEC E. coli translocated intimin receptor (Tir).
DE Tir: translocated intimin receptor; Hp90; enterohaemorrhagic;
KW EHEC; infection; diagnosis; vaccine.
XX OS Escherichia coli.
XX Key Location/Qualifiers
FH Misc-difference 453
FT /note= "encoded by codon of 1 apparent nucleotide,
FT causing frameshift in the DNA sequence"
XX WO9924576-A1.
XX 20-MAY-1999.
XX 10-NOV-1998; 98WO-CA01042.
XX 12-NOV-1997; 97US-0065130.
XX (UVER-) UNIV BRITISH COLUMBIA.
XX Devinney R, Finlay BB, Kenny B, Stein M;
XX WPI; 1999-337712/28.
DR N-PSDB; AAX58859.

XX New translocated intimin receptor useful for treating infection by
PT enteropathogenic or enterohemorrhagic Escherichia coli
XX
PS Claim 7; Page 55-58; 91pp; English.
XX
CC The present sequence represents Tir, a novel translocated intimin
CC receptor (formerly termed Hp90) from an enterohemorrhagic
CC Escherichia coli (EHEC) strain. The sequence was deduced from an
CC isolated tir polynucleotide (see AAX58859). Tir proteins are
CC secreted by attaching and effacing pathogens such as EHEC and EPEC
CC (see AAY06220) E. coli. The bacterial pathogens insert their own
CC receptors into mammalian cell surfaces, to which the pathogen then
CC adheres to trigger additional host signaling events and actin
CC nucleation. Diagnosis of disease caused by pathogenic E. coli can
CC be performed by use of antibodies that bind to Tir to detect the
CC protein or the use of nucleic acid probes for detection of nucleic
CC acids encoding Tir polypeptide. Isolated Tir nucleic acids, Tir
CC peptides, a recombinant method for producing recombinant Tir,
CC antibodies which bind to Tir, and a kit for the detection of
CC Tir-producing E. coli are provided. A method of immunising a host
CC with Tir to induce a protective immune response is also provided.
CC In addition, Tir fusion proteins can be used in attenuated E. coli
CC to induce a cell-mediated immune response to other polypeptides,
CC e.g. antigens. A method for screening for compounds which
CC interfere with the binding of bacterial pathogens to their
CC receptors is further provided.

XX Sequence 559 AA;

Query Match 100.0%; Score 2851; DB 20; Length 559;
Best Local Similarity 100.0%; Pred. No. 9.3e-213;
Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPIGNLGHNPVNNNSIPPAPLPSTQDAGGGRGQLINSTGLGRALFTPVNRNMAVSGD 60
DB 1 mpignlghnpvnnnsippaplpstqdaggrgqlinstglgralftpvnrnmasdsgd 60

QY 61 NPASDVPGLVPNMPRLAASEITLNDGFVLDHDPGLDPLNRQIGSSVFRVETOEDGKHIA 120
DB 61 nrasdvpglvpnmprlaasetlndgfvldhdpdpldnrlqigssvfrvetqdgkhia 120

QY 121 VQGRNGVTSVVLSDQEVARLQSDPEKDKRFFVTGGRGAGHAMVTVASDITEARQRI 180
DB 121 vqgrngvtsvvlsgdevarlqsdpekdkrffvtggrgaghamvtvasditearqri 180

QY 181 ELLEPKGTGESKAGESKVGELRESNSGAENTTQTSTSTSSLRSDPKLWALGTVAT 240
DB 181 ellepkgtgeskageskvgelesnsngaenttqtststsslrspkylwalgvtat 240

QY 241 GLIGLAATGIVCALATPPDSPTTTPDAAASATETATRODLTKFAFONPDNQKVNIDE 300
DB 241 gliglaatgivalatppdsptttdpaaasatetatrdoltkeafonpdnqkvnde 300

QY 301 LGNAIPSGVLKDDVVANIEQAKAGEAKQQA IENNAQAQKYYDEQAKROEELKVSSG 360
DB 301 lgnai-psgvlkddvvanieeqakageakqqa iennaaqaqkyydeqakroeelkvssg 360

QY 361 AGYGLSGALILGGGIVAVTAALHKKRNPQVEQTITTTTTTTSARTVENKPANNTPAQG 420
DB 361 agyglsgalilgggivaavtaalhkkrn-pqveqtitTTTTTTSartvenkpanntpaqg 420

QY 421 NVDTPGSEDTMSSRSSMASTSTFTFDTSSIGGPCRIRMLKHKMRRCLLILRLF 480
DB 421 nvdtpgsedtmssrsmaststftfDTSSIGGPCRIRMLKHKMRRCLLILrlf 480

QY 481 RIWGIQISVYSTIQHPRRDTTNDGARLGNPSAGTQSTYARLALSGGLRHDMMGLTGGS 540
DB 481 riwgiqisvystiqhprdrdtTNDGARLGNpsagTQSTyARLalSGGLrHdmmglTggs 540

QY 541 NSAVTNSNPPAGSHRFV 559
DB 541 nsavtNSNPPAGshrfv 559

Db 541 nsavtNSNPPAGSHRFV 559

RESULT 2
AAY06220
ID AAY06220 standard; Protein; 549 AA.
XX
AC AAY06220;
XX
DT 16-AUG-1999 (first entry)
XX
DE EPEC E. coli translocated intimin receptor (Tir).
XX
KW Tir; translocated intimin receptor; Hp90; enteropathogenic;
XX EPEC; infection; diagnosis; vaccine.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT Domain 234..253
FT /note= "putative transmembrane domain"
FT Domain 364..386
FT /note= "putative transmembrane domain"
FT Misc-difference 180
FT /note= "encoded by AAA"
FT Misc-difference 314
FT /note= "given as Xaa in the specification; Lys
FT is deduced from the DNA sequence"
XX
XX WO9924576-A1.
XX
XX 20-MAY-1999.
XX
XX 10-NOV-1998; 98WO-CA01042.
XX
XX 12-NOV-1997; 97US-0065130.
XX
XX (UYBR-) UNIV BRITISH COLUMBIA.
XX
XX Devlinney R, Finlay BB, Kenny B, Stein M;
XX
XX WPI; 1999-337712/28.
XX
XX N-PSDB; AAX58859.
XX
XX New translocated intimin receptor useful for treating infection by
XX enteropathogenic or enterohemorrhagic Escherichia coli
XX
XX Claim 6; Page 55-58; 91pp; English.
XX
XX The present sequence represents Tir, a novel translocated intimin
XX receptor (formerly termed Hp90) from an enteropathogenic
XX Escherichia coli (EPEC) strain. The sequence was deduced from an
XX isolated tir polynucleotide (see AAX58858). Tir proteins are
XX secreted by attaching and effacing pathogens such as EPEC and EHEC
XX (see AAY06221) E. coli. The bacterial pathogens insert their own
XX receptors into mammalian cell surfaces, to which the pathogen then
XX adheres to trigger additional host signaling events and actin
XX nucleation. Diagnosis of disease caused by pathogenic E. coli can
XX be performed by use of antibodies that bind to Tir to detect the
XX protein or the use of nucleic acid probes for detection of nucleic
XX acids encoding Tir polypeptide. Isolated Tir nucleic acids, Tir
XX peptides, a recombinant method for producing recombinant Tir,
XX antibodies which bind to Tir, and a kit for the detection of
XX Tir-producing E. coli are provided. A method of immunising a host
XX with Tir to induce a protective immune response is also provided.
XX In addition, Tir fusion proteins can be used in attenuated E. coli
XX to induce a cell-mediated immune response to other polypeptides,
XX e.g. antigens. A method for screening for compounds which
XX interfere with the binding of bacterial pathogens to their
XX receptors is further provided.

XX Sequence 549 AA;

Query Match 52.5%; Score 1498; DB 20; Length 549;
Best Local Similarity 54.8%; Pred. No. 6.6e-108;
Matches 323; Conservative 64; Mismatches 132; Indels 70; Gaps 12;
QY 1 MPIGNLHNPVNNISPPAPLPSTQDGA--GGRQLINSTGPLGSRALFTPVVRNSMADS 58
DB 1 mpignlgnvngnhlpppplpstqdaargtgthllsstgalgsralfplrnsmads 60
QY 59 GNRASDVPGLVPNPMRLAA--SETLNDGEVLVHDGPLETLNRQIGSSVFRVETQDEG 116
DB 61 vdsr--dipgiptnpsrlaatsetcllgfevlhdkpdlidltqipsatrfvqvadg 118
QY 117 KHIAVGQRNGVETSIVLSDQEVARLQSIDPECKDFVETGGRGGAGHAMVVASDITEAR 176
DB 119 thaaigeKnglevsvtlspqewsslqsdtegnrfvftggrgshgmvtvasdiaear 178
QY 177 QRILELPEKGTG-----ESKGAGESKGVGELRESNCAENTTQTSTSTSLRS 227
DB 179 trilakldpnhggrrpkdvdtrsvvgvsagi-----ddgv--vsethtsttnsvrs 230
QY 228 DPKLWALGTVATGLIGLAAGIVQALALTPPDSPPTTDPDAAASATETATRDQLTKEA 287
DB 231 dpkfwsvgalaaglaglaatgiaqalaltpepdpttdpdqaanaaesatkdlqtgea 290
QY 288 FQNPQKXVNIDELGNATPSGVKDDVANTIEEQAKAAGEEAKQAIENNAQAOKKYDEQ 347
DB 291 fknpenkvndangnalpsgeikddiveqlaqakeagevarqgavesnaqagryedq 350
QY 348 QAKQOEELKVSAGYGLSGALILGGIGVAVTAALHRKNQRPVEQTTTTTTTTTSART 407
DB 351 harrqeelqlesgiyglssalivaggiagvtalhrnpaetttttt-----ht 403
QY 408 VENKPAANTPAQNVDTPCSEDTEMRSSMASTSTFFDSS-----IGGPCRIR 458
DB 404 vvqqgtgglpqhkvampqerrifsdrrdsgsvasthwsdsssevvnpyaevvg---ar 460
QY 459 MMLKHCRCRCLLILIRLFRWIGIISVYVSTIQHPP-----RDTDNG---ARLLG 510
DB 461 nsishq-----peehlydevaadpyvsiqnfsg99gvtgrlig 500
QY 511 NPSAGIQSTYARLALSGRLDMGLTGGSNAVNTSNPPAPGSHRFV 559
DB 501 tpgqglqstyalansgllrgmgltsggetavsvnaaptgprfv 549

RESULT 3
AAB20576
ID AAB20576 standard; protein; 107 AA.
XX AC AAB20576;
XX DT 08-DEC-2000 (first entry)
XX DE Intimin C-terminal Tir binding domain amino acid sequence.
XX KW Intimin; Tir binding domain; inhibitor; intimin adhesion; screening;
XX KW Tir-independent eukaryotic cell binding activity; bacterial infection;
XX KW diarrhoea; antibacterial.
XX OS Unidentified.
XX PN WO200045173-A1.
XX PD 03-AUG-2000.
XX PF 31-JAN-2000; 2000WO-GB00254.
XX PR 29-JAN-1999; 99GB-0001897.
XX PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX PI Frankel GM, Matthews SJ, Hale CB, Dougan G;

XX WPI; 2000-499357/44.
XX DR Screening for inhibitors of intimin binding to eukaryotic cells, for
XX PT use in diagnosing, preventing and treating bacterial infections,
XX PT especially *Escherichia coli* O157:H7 -
XX PS Claim 8; Page 76; 96pp; English.
XX CC The present invention describes a method of screening for an inhibitor
XX CC of intimin binding to eukaryotic cells. The method comprises exposing an
XX CC intimin polypeptide having a Tir-independent cell binding activity to
XX CC test agents, and obtaining an inhibitor based on its ability to bind the
XX CC polypeptide. The inhibitors are used in the prevention, treatment and/or
XX CC diagnosis of bacterial infections, preferably by enteropathic and/or
XX CC enterohaemorrhagic *Escherichia coli*, Shiga toxinigenic *E. coli*, *Hafnia*
XX CC *alvei* or *Citrobacter freundii*, or especially *E. coli* O157:H7. The
XX CC infections cause a histopathological effect known as attachment and
XX CC effacement on intestinal epithelial cells. The inhibitors can be used
XX CC to produce food supplements or additives, especially where the food is
XX CC a milk substitute. The method can be used to sort cells based on their
XX CC ability to bind to a Tir independent cell binding domain of an intimin
XX CC polypeptide. Polypeptides having Tir-independent intimin binding
XX CC activity can be used to produce a vaccine against a bacterial disease.
XX CC The present sequence represents a specifically claimed intimin C-terminal
XX CC Tir binding domain amino acid sequence, for use in the method of the
XX CC present invention.
XX SQ Sequence 107 AA;
Query Match 14.2%; Score 406; DB 21; Length 107;
Best Local Similarity 72.0%; Pred. No. 2.9e-24;
Matches 77; Conservative 18; Mismatches 12; Indels 0; Gaps 0;
QY 252 QALALTPPDSPPTTDPDAAASATETATRDQLTKEAFQNPDKVNIDELGNATPSGVLK 311
DB 1 qalaltpepdpttdpdqaanaaesatkdlqtgeafknpenkvndangnaipsgelk 60
QY 312 DDVANTIEEQAKAAGEEAKQAIENNAQAOKKYDEQAKRQELKVS 358
DB 61 ddiveqlaqakeagevarqgavesnaqagryedqharrqeelqls 107
RESULT 4
AAY99408
ID AAY99408 standard; Protein; 596 AA.
XX AC AAY99408;
XX DT 08-AUG-2000 (first entry)
XX DE Human PRO1342 (UNQ697) amino acid sequence SEQ ID NO:243.
XX KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
XX KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX OS Homo sapiens.
XX PN WO200012708-A2.
XX PD 09-MAR-2000.
XX PF 01-SEP-1999; 99WO-US20111.
XX PR 01-SEP-1998; 98US-0098716.
XX PR 01-SEP-1998; 98US-0098749.
XX PR 01-SEP-1998; 98US-0098750.
XX PR 02-SEP-1998; 98US-0098803.
XX PR 02-SEP-1998; 98US-0098821.
XX PR 02-SEP-1998; 98US-0098843.
XX PR 09-SEP-1998; 98US-0099536.
XX PR 09-SEP-1998; 98US-0099596.

| | | | | | |
|----|--------------|---------------|----|--------------|---|
| PR | 09-SEP-1998; | 98US-0099598. | PR | 22-OCT-1998; | 98US-0105266. |
| PR | 09-SEP-1998; | 98US-0099602. | PR | 26-OCT-1998; | 98US-0105693. |
| PR | 09-SEP-1998; | 98US-0099642. | PR | 26-OCT-1998; | 98US-0105694. |
| PR | 10-SEP-1998; | 98US-0099741. | PR | 27-OCT-1998; | 98US-0105807. |
| PR | 10-SEP-1998; | 98US-0099754. | PR | 27-OCT-1998; | 98US-0105881. |
| PR | 10-SEP-1998; | 98US-0099763. | PR | 27-OCT-1998; | 98US-0105882. |
| PR | 10-SEP-1998; | 98US-0099792. | PR | 27-OCT-1998; | 98US-0106062. |
| PR | 10-SEP-1998; | 98US-0099808. | PR | 28-OCT-1998; | 98US-0106023. |
| PR | 10-SEP-1998; | 98US-0099812. | PR | 28-OCT-1998; | 98US-0106029. |
| PR | 10-SEP-1998; | 98US-0099815. | PR | 28-OCT-1998; | 98US-0106030. |
| PR | 10-SEP-1998; | 98US-0099816. | PR | 28-OCT-1998; | 98US-0106032. |
| PR | 15-SEP-1998; | 98US-0100385. | PR | 28-OCT-1998; | 98US-0106033. |
| PR | 15-SEP-1998; | 98US-0100388. | PR | 28-OCT-1998; | 98US-0106178. |
| PR | 15-SEP-1998; | 98US-0100390. | PR | 28-OCT-1998; | 98US-0106248. |
| PR | 16-SEP-1998; | 98US-0100384. | PR | 29-OCT-1998; | 98US-0106384. |
| PR | 16-SEP-1998; | 98US-0100627. | PR | 29-OCT-1998; | 98US-0106384. |
| PR | 16-SEP-1998; | 98US-0100661. | PR | 29-OCT-1998; | 98US-0108500. |
| PR | 16-SEP-1998; | 98US-0100662. | PR | 30-OCT-1998; | 98US-0106464. |
| PR | 16-SEP-1998; | 98US-0100664. | PR | 30-OCT-1998; | 98US-0106856. |
| PR | 17-SEP-1998; | 98US-0100663. | PR | 03-NOV-1998; | 98US-0106856. |
| PR | 17-SEP-1998; | 98US-0100684. | PR | 03-NOV-1998; | 98US-0106902. |
| PR | 17-SEP-1998; | 98US-0100710. | PR | 03-NOV-1998; | 98US-0106905. |
| PR | 17-SEP-1998; | 98US-0100711. | PR | 03-NOV-1998; | 98US-0106919. |
| PR | 17-SEP-1998; | 98US-0100919. | PR | 03-NOV-1998; | 98US-0106932. |
| PR | 17-SEP-1998; | 98US-0100930. | PR | 03-NOV-1998; | 98US-0106934. |
| PR | 18-SEP-1998; | 98US-0100848. | PR | 10-NOV-1998; | 98US-0107783. |
| PR | 18-SEP-1998; | 98US-0100849. | PR | 17-NOV-1998; | 98US-0108775. |
| PR | 18-SEP-1998; | 98US-0101014. | PR | 17-NOV-1998; | 98US-0108779. |
| PR | 18-SEP-1998; | 98US-0101068. | PR | 17-NOV-1998; | 98US-0108787. |
| PR | 18-SEP-1998; | 98US-0101071. | PR | 17-NOV-1998; | 98US-0108788. |
| PR | 22-SEP-1998; | 98US-0101279. | PR | 17-NOV-1998; | 98US-0108801. |
| PR | 23-SEP-1998; | 98US-0101471. | PR | 17-NOV-1998; | 98US-0108802. |
| PR | 23-SEP-1998; | 98US-0101472. | PR | 17-NOV-1998; | 98US-0108806. |
| PR | 23-SEP-1998; | 98US-0101473. | PR | 17-NOV-1998; | 98US-0108807. |
| PR | 23-SEP-1998; | 98US-0101475. | PR | 17-NOV-1998; | 98US-0108867. |
| PR | 23-SEP-1998; | 98US-0101476. | PR | 17-NOV-1998; | 98US-0108925. |
| PR | 23-SEP-1998; | 98US-0101477. | PR | 18-NOV-1998; | 98US-0108848. |
| PR | 23-SEP-1998; | 98US-0101479. | PR | 18-NOV-1998; | 98US-0108849. |
| PR | 24-SEP-1998; | 98US-0101738. | PR | 18-NOV-1998; | 98US-0108850. |
| PR | 24-SEP-1998; | 98US-0101741. | PR | 18-NOV-1998; | 98US-0108851. |
| PR | 24-SEP-1998; | 98US-0101743. | PR | 18-NOV-1998; | 98US-0108852. |
| PR | 24-SEP-1998; | 98US-0101915. | PR | 18-NOV-1998; | 98US-0108858. |
| PR | 24-SEP-1998; | 98US-0101916. | PR | 18-NOV-1998; | 98US-0108904. |
| PR | 29-SEP-1998; | 98US-0102207. | XX | | |
| PR | 29-SEP-1998; | 98US-0102240. | XX | | |
| PR | 29-SEP-1998; | 98US-0102307. | XX | | |
| PR | 29-SEP-1998; | 98US-0102330. | XX | | |
| PR | 29-SEP-1998; | 98US-0102331. | XX | | |
| PR | 30-SEP-1998; | 98US-0102484. | XX | | |
| PR | 30-SEP-1998; | 98US-0102487. | XX | | |
| PR | 30-SEP-1998; | 98US-0102570. | XX | | |
| PR | 30-SEP-1998; | 98US-0102571. | XX | | |
| PR | 01-OCT-1998; | 98US-0102684. | XX | | |
| PR | 01-OCT-1998; | 98US-0102687. | XX | | |
| PR | 02-OCT-1998; | 98US-0102965. | XX | | |
| PR | 06-OCT-1998; | 98US-0103258. | XX | | |
| PR | 06-OCT-1998; | 98US-0103449. | XX | | |
| PR | 07-OCT-1998; | 98US-0103314. | XX | | |
| PR | 07-OCT-1998; | 98US-0103315. | XX | | |
| PR | 07-OCT-1998; | 98US-0103328. | XX | | |
| PR | 07-OCT-1998; | 98US-0103395. | XX | | |
| PR | 07-OCT-1998; | 98US-0103396. | XX | | |
| PR | 07-OCT-1998; | 98US-0103401. | XX | | |
| PR | 08-OCT-1998; | 98US-0103633. | XX | | |
| PR | 08-OCT-1998; | 98US-0103678. | XX | | |
| PR | 08-OCT-1998; | 98US-0103679. | XX | | |
| PR | 14-OCT-1998; | 98US-0103711. | XX | | |
| PR | 20-OCT-1998; | 98US-0104257. | XX | | |
| PR | 20-OCT-1998; | 98US-0104987. | XX | | |
| PR | 20-OCT-1998; | 98US-0105000. | XX | | |
| PR | 20-OCT-1998; | 98US-0105002. | XX | | |
| PR | 21-OCT-1998; | 98US-0105104. | XX | | |
| PR | 22-OCT-1998; | 98US-0105169. | XX | | |
| | | | | | (GETH) GENENTECH INC. |
| | | | | | Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI; |
| | | | | | WPI; 2000-237871/20. |
| | | | | | N-PSDB; AAA37090. |
| | | | | | New mammalian DNA sequences encoding transmembrane, receptor or |
| | | | | | secreted PRO polypeptides, useful for screening of potential peptide or |
| | | | | | small molecule inhibitors of the relevant receptor/ligand interactions |
| | | | | | Claim 12; Fig 138; 773pp; English. |
| | | | | | AAA37022 to AAA37144 encode the new isolated human transmembrane, |
| | | | | | receptor or secreted PRO polypeptides given in AA99340 to AA99462. The |
| | | | | | transmembrane and receptor PRO proteins can be used for screening of |
| | | | | | potential peptide or small molecule inhibitors of the relevant |
| | | | | | receptor/ligand interactions. The polypeptides and nucleotide sequences |
| | | | | | encoding then have various industrial applications, including uses as |
| | | | | | pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent |
| | | | | | PCR primers and hybridisation probes used in the isolation of the PRO |
| | | | | | polypeptides from the present invention. |
| | | | | | Sequence 596 AA; |
| | | | | | Query Match 5.2%; Score 148.5; DB 21; Length 596; |
| | | | | | Best Local Similarity 17.7%; Pred No. 0.0029; |
| | | | | | Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19; |
| | | | | | QY 4 GNLGHPNVNNSIPPAPLPSTQDGAGGRQLNSTGPIGSRALFTPVNSMADSGDNA 63 |

```
Db 13 gilllleatns-----netsantgssvissg-----astatnsg 49
QY 64 SDVPGLPVNPMLAASEITLNDGFVLDHDPGLDLNRLQIGSSV-FRVETQEDGKHIAVG 122
Db 50 ssvtssgsvstatnsgsvtsn-gvsiv-tnsefhtstgstatnsefstassgisiatn 107
QY 123 QRNGVETS--VVLSDQEVARLQSIDPEGKDFVFTGGRGGAGHAMVTVASDITEARQRL 180
Db 108 sessttsgastatnse-----sstpsgastvtnsgsvtsgastatnseestvssras 163
QY 181 ELLEPKGTGESKAGESKGVGELRESNGAENTTQTSTSTSSRLSDPKLWALGTAVT 240
Db 164 tatnsestlsgastatn-sdssttsgastatnsesttsg-----astat 211
QY 241 GLIGLAATGIQVALALTPEDSPPTTDPDAAASATETATRDQLTKAEAFONPNQKNIDE 300
Db 212 n-----sesstvsrastatnseestt---sagastatnsesttngagatnsestts 264
QY 301 LG-----NAIPSGVLKDDVAVANIEEQAKAAGEAKQQAENNAQAOKKYDQQAQR 351
Db 265 sgastatnsestvsaga-----statnsesttsgast-----atn 302
QY 352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTSARTVEN 410
Db 303 sessttsgastatnsesttsgastatnsestsvsgistvtnsestpsgantatn 362
QY 411 KPANTPAQGNVDTPGSEDTMESRRSSMASTSTFTDSSIGGPCRIRMLMLKHCMIIR 470
Db 363 sessttsgantatnsestsvsgastatnsest---ttsg----- 401
QY 471 CRLILILRIFRIWGIQISVVYSTIOHPPRDITDNGARLLGNPSAGIQSTYARLA----- 524
Db 402 -----vstatnsesttsgastatnsesttsgastatnsesttsgastatnsest 441
QY 525 -LSGGL-----RHDGGLTGGSNSAVNTSNPPAPGS 555
Db 442 tvssgistvtnsesttsgantatnsgsvtsags 477

RESULT 5
AAB87575
ID AAB87575 standard; Protein; 596 AA.
AC AAB87575;
XX
XX 15-MAY-2001 (first entry)
DE Human PRO1342.
KW Human; PRO protein; mapping.
OS Homo sapiens.
XX
XX WO200116318-A2.
PD 08-MAR-2001.
PF 24-AUG-2000; 2000WO-US23328.
XX
XX 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 07-DEC-1999; 99US-0169495.
PR 09-DEC-1999; 99US-0170262.
PR 11-JAN-2000; 2000US-0175481.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 01-MAR-2000; 2000WO-US05601.
PR 03-MAR-2000; 2000US-0187202.
PR 25-APR-2000; 2000US-0199397.
PR 22-MAY-2000; 2000WO-US14042.
PR 05-JUN-2000; 2000US-0209832.
```

```
XX
PA (GETH ) GENENTECH INC.
XX
PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX
XX WPI; 2001-183260/18.
DR N-PSDB; AAF92107.
XX
XX Eighty four nucleic acids encoding PRO polypeptides, useful in
PT molecular biology, including use as hybridization probes, and in
PT chromosome and gene mapping.
XX
PS Claim 12; Fig 100; 278pp; English.
XX
CC The present sequence is a human PRO polypeptide (secreted and
CC transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC anti-PRO antibodies are useful for preparation of a medicament useful in
CC the treatment of a condition which is responsive to the PRO protein,
CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC employed as molecular weight markers for protein electrophoresis. The PRO
CC coding sequence has applications in molecular biology, including use as
CC hybridisation probes, and in chromosome and gene mapping.
XX
SQ Sequence 596 AA;
```

Query Match 5.2%; Score 148.5; DB 22; Length 596;

Best Local Similarity 17.7%; Pred. No. 0.0029;

Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

```
QY 4 GNIGHNPVNNISIPAPLPSPOTDGGGRGQLINSTGPLGSRALFTPVNSMADSGDRA 63
Db 13 gilllleatns-----netsantgssvissg-----astatnsg 49
QY 64 SDVPGLPVNPMLAASEITLNDGFVLDHDPGLDLNRLQIGSSV-FRVETQEDGKHIAVG 122
Db 50 ssvtssgsvstatnsgsvtsn-gvsiv-tnsefhtstgstatnsefstassgisiatn 107
QY 123 QRNGVETS--VVLSDQEVARLQSIDPEGKDFVFTGGRGGAGHAMVTVASDITEARQRL 180
Db 108 sessttsgastatnse-----sstpsgastvtnsgsvtsgastatnseestvssras 163
QY 181 ELLEPKGTGESKAGESKGVGELRESNGAENTTQTSTSTSSRLSDPKLWALGTAVT 240
Db 164 tatnsestlsgastatn-sdssttsgastatnsesttsg-----astat 211
QY 241 GLIGLAATGIQVALALTPEDSPPTTDPDAAASATETATRDQLTKAEAFONPNQKNIDE 300
Db 212 n-----sesstvsrastatnseestt---sagastatnsesttngagatnsestts 264
QY 301 LG-----NAIPSGVLKDDVAVANIEEQAKAAGEAKQQAENNAQAOKKYDQQAQR 351
Db 265 sgastatnsestvsaga-----statnsesttsgast-----atn 302
QY 352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTSARTVEN 410
Db 303 sessttsgastatnsesttsgastatnsestsvsgistvtnsestpsgantatn 362
QY 411 KPANTPAQGNVDTPGSEDTMESRRSSMASTSTFTDSSIGGPCRIRMLMLKHCMIIR 470
Db 363 sessttsgantatnsestsvsgastatnsest---ttsg----- 401
QY 471 CRLILILRIFRIWGIQISVVYSTIOHPPRDITDNGARLLGNPSAGIQSTYARLA----- 524
Db 402 -----vstatnsesttsgastatnsesttsgastatnsesttsgastatnsest 441
QY 525 -LSGGL-----RHDGGLTGGSNSAVNTSNPPAPGS 555
Db 442 tvssgistvtnsesttsgantatnsgsvtsags 477
```

RESULT 6

AAB66157
ID AAB66157 standard; protein; 596 AA.

XX AAB66157;

DT 02-APR-2001 (first entry)

XX Protein of the invention #69.

DE Secreted; transmembrane; gene therapy.

XX Unidentified.

OS WO200078961-A1.

PN 28-DEC-2000.

XX 18-FEB-2000; 2000WO-US04342.

XX 23-JUN-1999; 99US-0141037.

PR 20-JUL-1999; 99US-0144758.

PR 26-JUL-1999; 99US-0145698.

PR 01-SEP-1999; 99WO-US20111.

PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.

PR 02-DEC-1999; 99WO-US28551.

PR 16-DEC-1999; 99WO-US30095.

PR 05-JAN-2000; 2000WO-US00219.

PR 06-JAN-2000; 2000WO-US00376.

XX (GETH) GENENTECH INC.

PI Baker KP, Botstein D, Desnovers L, Eaton DL, Ferrara N, Fong S;
PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Pan J, Faoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
PI Watanabe CK, Williams PM, Wood WI;

XX WPI: 2001-071395/08.

DR Secreted and transmembrane proteins and nucleic acids designated PRO,
PT useful as hybridization probes, in chromosome and gene mapping and gene
PT therapy -
XX Claim 1; Fig 138; 787pp; English.

XX The present invention relates to secreted and transmembrane proteins.
XX These proteins and the DNA encoding them may be used as hybridization
CC probes, in chromosome and gene mapping and in the generation of
CC anti-sense RNA and DNA. They may also be used to generate either
CC transgenic animals or knockout animals which are in turn useful for
CC development and screening of therapeutically useful reagents.
CC The nucleic acids may also be used in gene therapy.

XX Sequence 596 AA;

Query Match 5.2%; Score 148.5; DB 22; Length 596;
Best Local Similarity 17.7%; Pred. NO. 0.0029;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

QY 4 GNIGHNPVNNIPAPPPLPSTDTGAGGQGLNSTGPLGSRALFTPVNNSMADSDNRA 63
DB 13 gillhleaatns-----netstnsantgssvissg-----astatsng 49

QY 64 SDVPGLPVNPMRLAASEITLNDGFVLDHDPDLTLNRQIGSSV-FRVETQDGRKHAIVG 122
DB 50 svtstsgvstatisgsvtsn-gvsiv-tnsfhttsstgstatnsfstatnsstgsviatn 107

QY 123 QRNGVETS--VVLSDQGYARLOSIDPEGKDKFVFCGGRGAGHAMVTVASDITEARQRL 180
DB 108 sessttsgastatnse-----sstpsgastvtnsgsvtsgastatnsesstvsrras 163

QY 181 ELLEPKGTGSKAGESKGVGELRESNCAENTEQTSTSTSLRSDPKLWALGTVAT 240

DB 164 tatsesstlsgastatn-sdssttsgastatnsessttsg-----astat 211
QY 241 GLIGLAATGIQALALTPEDSPPTTDPDAASATETATRDQLTKEAFQNDPNQKVNI 300
DB 212 n-----sesstvsrastatnsesstt---ssgastatnsesrttsngagatnsesstts 264
QY 301 LG-----NAIPSGVLKDDVANIIEQAKAAGEEAKQQAENNAQAQKYDQQA 351
DB 265 sgastatnsdsstvsnga---statnsessttsgast-----atn 302
QY 352 QBELKVSSGAGYGL-SGALLGGIGVAVTAALHRKNQVPEQTTTTTTTTTTTSARTVEN 410
DB 303 sessttsgastatnsdssttsgagatnsesstvsstgsvtstnsesstpsgagatn 362
QY 411 KPANTPAGQNVDTPGSEDTMESRRSSMASTSTSTFTDTSIGGPCIRMLMLKHCMI 470
DB 363 sessttsgastatnsesstvsstgastatnses-----ttssg----- 401
QY 471 CRLLILIRLFRWIGIOISVYSTIQHPPTDNTDNGARLLGNPSAGIOSYARLA----- 524
DB 402 -----vstatnsessttsgastatnsdssttsgastatnses 441
QY 525 -LSGGL-----RHDGMGLTGSNSAVNTSNPPAPGS 555
DB 442 tvsgstvtnsessttsgastatnsesstvsstgsvtsags 477

RESULT 7

AAR10562
ID AAR10562 standard; Protein; 1959 AA.

XX AAR10562;

DT 15-APR-1991 (first entry)

XX Mutant protease (delta137-139).

XX Mutant protease gene; fermentation; foodstuff; flavouring;
KW Lactic acid bacteria.

XX Lactococcus lactis SK11.

PH Key Location/Qualifiers

FT Peptide 1..187 /label= signal_peptide

FT Protein 188..1959 /label= mature_peptide

XX EP411715-A.

XX 06-FEB-1991.

XX 02-AUG-1990; 90EP-0202113.

XX 04-AUG-1989; 89NL-0002010.

XX (NEZU-) NED INST ZUIVELONDE.

XX Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;

XX WPI: 1991-038622/06.

XX N-PSDB; AAQ10416.

XX Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings

PS Disclosure; Fig 1(1-7)+5(b); 29 pp; English.

XX The mutant protease having new cleavage specificities is

CC obtained by deleting three amino acids.

CC This mutant may then be used to prepare hybrid proteases,

CC The mutant protease H1570/K1500 having an amino acid

CC obtained by replacing two amino acids.
CC This mutant may then be used to prepare hybrid proteases.
CC the fusion being between a type I and a type III protease of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAQ10411-17 and AAQ10870-71.
XX
SQ Sequence 1962 AA;

Query Match 5.0%; Score 141.5; DB 12; Length 1962;
Best Local Similarity 22.8%; Pred. No. 0.051;
Matches 114; Conservative 61; Mismatches 211; Indels 113; Gaps 20;

QY 36 INSTGSLGSRALFTVPR-----NSMADSGDNRAVDVPGLPVPMR-LAASEITLND 85
DB 1452 insgkp-ghmaidqpvklegknvltavtseadnttktvyyepkktlaaptvps- 1509
QY 86 GPEVLHDHGPLDNLNRQIGSSVFRVETQDGHIAVGQRNGVETSVVLSDOEYARLQSID 145
DB 1510 -----ttepaqvtltanaaa-----tgetvqysadgktyqdv- -----aagvtit 1551
QY 146 PEGKDKFVTGGRGAGHAMVTVASDI-----TEARQRIELLEPKGTGCKESGAGES 197
DB 1552 angtfkfstldlygnespavdyvvtvnikadpqaqlaakqeltnliasaktilsasgkydd 1611
QY 198 KGVGLRESNSGAENTTQTSTSSSLR-SDPKLWLALGTVATGLIGLAATGIVOALAL 256
DB 1612 atttalaatqkaq-taldqtnasvdsitganrdlqtainqlaaklpadkktkslnlqlqs 1670
QY 257 TPE----PDSPTTTPDPAASATETATRDQLTKEA---FQNPQNKVNIDELGNAIPSGVL 310
DB 1671 vkdalgtldlgnqtdp--stgkfttaaldlvaqagagtqtdqqlqatlak-----i 1719
QY 311 KDDVVANIEEQAKAGE-----EAKQQA-----ENNAQAQKYDEQQAQR 351
DB 1720 ldevlakraegikaatpaevgnakdaatgktwyadiadltltsgqasadasdklahlqalq 1779
QY 352 QDELKVS-----GAGYGLSGALILGGGIGVAVTA-----ALHRKNOP 389
DB 1780 slktkvaaveaaktvgkgdtgtsdkggggtgtpapagdgkdkgdegspssggnlp 1839
QY 390 VEQTTTTTTTTTSAR-----TVENKPNANTPAQGNVD---TPGSEDVTMESRSSM- 438
DB 1840 tkpattsttdtdtdrngqltsdgtsgkkggggtgtpapagdgkdkgdegspssggnlp 1899
QY 439 ----ASTSTFTDTSIGG 453
DB 1900 tnpattsttdtdtdrng 1918

RESULT 11
AA10558
ID AAR10558 standard; Protein; 1962 AA.
XX
AC AAR10558;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease (A137G/K138P/W139P).
XX
KW Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT /label= signal_peptide

FT Protein 188..1962
FT /label= mature_protein
XX EP4111715-A.
XX 06-FEB-1991.
XX 02-AUG-1990; 90EP-02021113.
XX 04-AUG-1989; 89NL-0002010.
XX (NEZU-) NED INST ZUIVELONDE.
PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX WPI; 1991-038622/06.
DR N-PSDB; AAQ104112.
XX
PT Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
PS Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX
CC The mutant protease having new cleavage specificities is
CC obtained by replacing three amino acids.
CC This mutant may then be used to prepare hybrid proteases, of
CC the fusion being between a type I and a type III protease, of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAQ10411-17 and AAQ10870-71.
XX
SQ Sequence 1962 AA;

Query Match 5.0%; Score 141.5; DB 12; Length 1962;
Best Local Similarity 22.8%; Pred. No. 0.051;
Matches 114; Conservative 61; Mismatches 211; Indels 113; Gaps 20;

QY 36 INSTGSLGSRALFTVPR-----NSMADSGDNRAVDVPGLPVPMR-LAASEITLND 85
DB 1452 insgkp-ghmaidqpvklegknvltavtseadnttktvyyepkktlaaptvps- 1509
QY 86 GPEVLHDHGPLDNLNRQIGSSVFRVETQDGHIAVGQRNGVETSVVLSDOEYARLQSID 145
DB 1510 -----ttepaqvtltanaaa-----tgetvqysadgktyqdv- -----aagvtit 1551
QY 146 PEGKDKFVTGGRGAGHAMVTVASDI-----TEARQRIELLEPKGTGCKESGAGES 197
DB 1552 angtfkfstldlygnespavdyvvtvnikadpqaqlaakqeltnliasaktilsasgkydd 1611
QY 198 KGVGLRESNSGAENTTQTSTSSSLR-SDPKLWLALGTVATGLIGLAATGIVOALAL 256
DB 1612 atttalaatqkaq-taldqtnasvdsitganrdlqtainqlaaklpadkktkslnlqlqs 1670
QY 257 TPE----PDSPTTTPDPAASATETATRDQLTKEA---FQNPQNKVNIDELGNAIPSGVL 310
DB 1671 vkdalgtldlgnqtdp--stgkfttaaldlvaqagagtqtdqqlqatlak-----i 1719
QY 311 KDDVVANIEEQAKAGE-----EAKQQA-----ENNAQAQKYDEQQAQR 351
DB 1720 ldevlakraegikaatpaevgnakdaatgktwyadiadltltsgqasadasdklahlqalq 1779
QY 352 QDELKVS-----GAGYGLSGALILGGGIGVAVTA-----ALHRKNOP 389
DB 1780 slktkvaaveaaktvgkgdtgtsdkggggtgtpapagdgkdkgdegspssggnlp 1839
QY 390 VEQTTTTTTTTTSAR-----TVENKPNANTPAQGNVD---TPGSEDVTMESRSSM- 438
XX

```

Db 1840 tkpattsttdtdtrngltsdskgggqtpapagdigdkdgdsgsqssggnip 1899
QY 439 ----ASTSFTFTSSIGG 453
Db 1900 tnpattsttdtdtrng 1918

RESULT 12
AAR10559
ID AAR10559 standard; Protein; 1962 AA.
XX
AC AAR10559;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease (A137G/K138L/T139A).
XX
KW Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT Protein /label= signal_peptide
FT /label= mature_protein
XX
EP411715-A.
XX
PD 06-FEB-1991.
XX
PF 02-AUG-1990; 90EP-0202113.
XX
PR 04-AUG-1989; 89NL-0002010.
XX
PA (NEZU-) NED INST ZUIVELONDE.
XX
PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX
DR N-PSDB; AAQ10413.
XX
PT Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
PS Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX
CC The mutant protease having new cleavage specificities is
CC obtained by replacing three amino acids.
CC This mutant may then be used to prepare hybrid proteases,
CC the fusion being between a type I and a type III protease of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAQ10411-17 and AAQ10870-71.
XX
SQ Sequence 1962 AA;

Query Match 5.0%; Score 141.5; DB 12; Length 1962;
Best Local Similarity 22.8%; Pred. No. 0.051;
Matches 114; Conservative 61; Mismatches 211; Indels 113; Gaps 20;

QY 36 INSTGPIGSRALFTVPR-----NSMADSGDNRASDPVGLPVNPMR-LAASEITLND 85
Db 1452 insgkp-gmaidqpvkllgknvltavtdsedntttnitvyyepkktlaaptvtps- 1509
QY 86 GFEVLHDHGLDNLNRIGSSVFRVETQEDGKHIAVGQRNGVETSVLSQDEYARLQSID 145

```

```

Db 1510 -----ttepactvtltanaaa-----tgetvdsadgktyqdvpp-----aagvtit 1551
QY 146 PEGKDKFVFTGGGAGHAMVTVASDI-----TEARQRIELLEPKGTGESKAGES 197
Db 1552 angtfkfstldlygnespavdyvvttnikaddpaqiqaakqeltnliaasaktlsasgkydd 1611
QY 198 KGVGELRESNGAENTETQTSTSTSSLR-SDPKLWALGTVATGLIGLAATGIIVQALAL 256
Db 1612 atttalaatqkaq-caldqtnasvdsltganrldlqatnlaaklpadcktsllnqlqs 1670
QY 257 TPE---PDSPTTTPDPAASATETATPDOLTKEA---FONPDNQKVNIDELGNAIPSGVL 310
Db 1671 vkdalgtldlgnqtdp--stgkftfaaldldlvaqagsgttdqqlqatlak-----i 1719
QY 311 KDDVWANEIEQAKAAGE---EAKQQAII-----ENNAQAQKYDEQQAAR 351
Db 1720 ldevlaklaegikaatpaevgnakdaatgktyadiadtltsgasadasdklahlqaiq 1779
QY 352 QEELKVSS-----GAGYGLSGALILGGIGVAVTA-----ALHRKNQP 389
Db 1780 slktkvaavaeaaakvgkgdgtttsdkgggqtpapagdtgdkdgdsgsqssggnip 1839
QY 390 VEQTFTTFTTFTTTSAR-----TVENKFPANNTPAQGNVD---TPGSEDTMESRRSSM- 438
Db 1840 tkpattsttdtdtrngltsdskgggqtpapagdigdkdgdsgsqssggnip 1899
QY 439 ----ASTSFTFTSSIGG 453
Db 1900 tnpattsttdtdtrng 1918

RESULT 13
AAR10563
ID AAR10563 standard; Protein; 1962 AA.
XX
AC AAR10563;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease (K748T).
XX
KW Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT Protein /label= signal_peptide
FT /label= mature_protein
XX
EP411715-A.
XX
PD 06-FEB-1991.
XX
PF 02-AUG-1990; 90EP-0202113.
XX
PR 04-AUG-1989; 89NL-0002010.
XX
PA (NEZU-) NED INST ZUIVELONDE.
XX
PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX
DR N-PSDB; AAQ10417.
XX
PT Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
PS Disclosure; Fig 1(1-7)+5(a); 29 pp; English.

```


XX The mutant protease K748T having new cleavage specificities is
 CC obtained by carrying out single amino acid substitutions.
 CC This mutant may then be used to prepare hybrid proteases,
 CC the fusion being between a type I and a type III protease of
 CC L.lactis Wg2 and SK11.
 CC The product has modified properties, e.g. thermostability,
 CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
 CC compared to the parent protease(s). The proteases can be used for
 CC preparing products (butter cheese, human and animal foodstuffs)
 CC prepared with the aid of lactic acid bacteria.
 CC See also AAQ10411-17 and AAQ10870-71.
 XX
 SQ Sequence 1962 AA;

Query Match 5.0%; Score 141.5; DB 12; Length 1962;
 Best Local Similarity 22.8%; Pred. No. 0.051;
 Matches 114; Conservative 61; Mismatches 211; Indels 113; Gaps 20;

QY 36 INSTGPGSRALFTVPR-----NSMADSGDNRASDPVGLPVNPMR-LAASEITLND 85
 Db 1452 insgkp-ghmaidqpvklllegknvltavtdeedtttknitvyyepkktlaaptvtps- 1509
 QY 86 GFEVLHDHGFLDLNRQIGSSVFRVETQEDGKHIAVGQRNGVETSIVLSDDQYARLQSID 145
 Db 1510 -----ttepactvtltanaaa---tgetvqysadgktyqdv-----aagvtit 1551
 QY 146 PGKDKFVTGGRGGAGHAMVTVASDI-----TEARQRILELLEPKGTGESKAGES 197
 Db 1552 angtfkfstldlygnespavdyvvtnikaddpaqlqaakqeltnliasaaktsasgkydd 1611
 QY 198 KGVGELRESNSGAENTETQSTSTSSLR-SDPKLWALGTVATGLIGLAANGIVQALAL 256
 Db 1612 attalaaatqkaq-taldqtnasvdsldgandrdlqtainglaaklpadkksllnqlqs 1670
 QY 257 TPE---PDSPTTTPDAAASATETATRDOLTKEA---FQNPQNKVNIDELGNAPSGVL 310
 Db 1671 vkdalgtldlgnqtdp--sgkfttaaldvlvaqagagtqtdqlqatlak-----i 1719
 QY 311 KDDVVANIEEQAKAGE-----EAKQQA-----ENNAQAKKYDQEQAKR 351
 Db 1720 ldeviaklaegikaatpaevgnakdaatgktywadiadltltsgqasadasdklahlqalq 1779
 QY 352 QBELKVSS-----GAGYGLSGALILGGGIGVAVTA-----ALHRKNQP 389
 Db 1780 slktkvaavaeaaktvgkgdgttgsdkkggggtpapapgdgkdgdegspssggnip 1839
 QY 390 VEQTTTTTTTTTTTSAR-----TVENKPANNTPAQGNVD---TPGSDTMSRRSSM- 438
 Db 1840 tkpattsttdtdtdrngqltsdtkgsgggtpapapgdgkdgdegspssggnip 1899
 QY 439 ----ASTSTFTDTSIGG 453
 Db 1900 tnpattsttdtdtdrng 1918

RESULT 14
 AARI0941
 ID AARI0941 standard; Protein; 1968 AA.
 XX
 AC AARI0941;
 XX
 DT 15-APR-1991 (first entry)
 XX
 DE Mutant protease (delta137-139/ins9).
 XX
 KW Mutant protease gene; fermentation; foodstuff; flavouring;
 OS lactic acid bacteria.
 XX
 OS Lactococcus lactis SK11.
 XX
 FH Key Location/Qualifiers

FT Peptide 1..187
 FT Protein /label= sig_peptide
 FT 188..1968
 FT /label= mat_protein
 XX
 PN EP411715-A.
 XX
 PD 06-FEB-1991.
 XX
 PF 02-AUG-1990; 90EP-0202113.
 XX
 PR 04-AUG-1989; 89NL-0002010.
 XX
 PA (NEZU-) NED INST ZUIVELONDE.
 XX
 PI Vos PAJ, Sieren RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
 DR WPI; 1991-038622/06.
 DR N-PSDB; AAQ10871.
 XX
 PT Mutant protease gene(s) and protease(s) - derived from type I and
 PT III protease genes from lactococcal strains, used in fermentation
 PT foodstuffs and flavourings
 XX
 PS Disclosure; Fig 1(1-7)+5(b); 29pp; English.
 XX
 CC The wild-type L.lactis SK11 protease gene sequence was determined
 CC by the applicant (EP-307011).
 CC The mutant protease having new cleavage specificities is
 CC obtained by deleting three amino acids (nine bps) and inserting
 CC 9 other residues.
 CC This mutant may then be used to prepare hybrid proteases,
 CC the fusion being between a type I and a type III protease of
 CC L.lactis Wg2 and SK11.
 CC The product has modified properties, e.g. thermostability,
 CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
 CC compared to the parent protease(s). The proteases can be used for
 CC preparing products (butter cheese, human and animal foodstuffs)
 CC prepared with the aid of lactic acid bacteria.
 CC See also AAQ10411-17 and AAQ10870-71.
 XX
 SQ Sequence 1968 AA;

Query Match 5.0%; Score 141.5; DB 12; Length 1968;
 Best Local Similarity 22.8%; Pred. No. 0.052;
 Matches 114; Conservative 61; Mismatches 211; Indels 113; Gaps 20;

QY 36 INSTGPGSRALFTVPR-----NSMADSGDNRASDPVGLPVNPMR-LAASEITLND 85
 Db 1458 insgkp-ghmaidqpvklllegknvltavtdeedtttknitvyyepkktlaaptvtps- 1515
 QY 86 GFEVLHDHGFLDLNRQIGSSVFRVETQEDGKHIAVGQRNGVETSIVLSDDQYARLQSID 145
 Db 1516 -----ttepactvtltanaaa---tgetvqysadgktyqdv-----aagvtit 1557
 QY 146 PGKDKFVTGGRGGAGHAMVTVASDI-----TEARQRILELLEPKGTGESKAGES 197
 Db 1558 angtfkfstldlygnespavdyvvtnikaddpaqlqaakqeltnliasaaktsasgkydd 1617
 QY 198 KGVGELRESNSGAENTETQSTSTSSLR-SDPKLWALGTVATGLIGLAANGIVQALAL 256
 Db 1618 attalaaatqkaq-taldqtnasvdsldgandrdlqtainglaaklpadkksllnqlqs 1676
 QY 257 TPE---PDSPTTTPDAAASATETATRDOLTKEA---FQNPQNKVNIDELGNAPSGVL 310
 Db 1677 vkdalgtldlgnqtdp--sgkfttaaldvlvaqagagtqtdqlqatlak-----i 1725
 QY 311 KDDVVANIEEQAKAGE-----EAKQQA-----ENNAQAKKYDQEQAKR 351
 Db 1726 ldeviaklaegikaatpaevgnakdaatgktywadiadltltsgqasadasdklahlqalq 1785
 QY 352 QBELKVSS-----GAGYGLSGALILGGGIGVAVTA-----ALHRKNQP 389

Db 1786 slkckvaavaeakvkgdgtgtsdkgggqgtpapapadtgdkdgdsgqpsgggnip 1845
QY 390 VEQTTTTTTTTTSAR-----TVENKPNANTPAQGNVD---TPGSEDTMESRRSSM- 438
Db 1846 tkpattstttdtdtrngqitsgtsdkgggqgtpapapgdigdkdgdsgqpsgggnip 1905
QY 439 ----ASTSSFFDFTSSIGG 453
Db 1906 tnpattstttdtdtdtrng 1924

RESULT 15
AAR10940
ID AAR10940 standard; Protein; 1974 AA.
XX
AC AAR10940;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease (delta137-139/ins15).
XX
KW Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT /label= SIG_PEPTIDE
FT Protein 188..1974
FT /label= MAT_PROTEIN
XX
PN EP411715-A.
XX
PD 06-FEB-1991.
XX
PF 02-AUG-1990; 90EP-0202113.
XX
PR 04-AUG-1989; 89NL-0002010.
XX
PA (NEZU-) NED INST ZUIVELONDE.
XX
PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX
DR WPI: 1991-038622/06.
DR N-PSDB; AAQ10870.
XX
PT Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
PS Disclosure; Fig 1(1-7)+5(b); 29pp ; English.
XX
CC The wild-type L.lactis SK11 protease gene sequence was determined
CC by the applicant (EP-307011).
CC The mutant protease having new cleavage specificities is
CC obtained by deleting three amino acids (nine bps) and inserting
CC 15 other residues.
CC This mutant may then be used to prepare hybrid proteases,
CC the fusion being between a type I and a type III protease of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAQ10411-17 and AAQ10870-71.
XX
SQ Sequence 1974 AA;

Query Match 5.0%; Score 141.5; DB 12; Length 1974;

Best Local Similarity 22.8%; Pred. No. 0.052;
Matches 114; Conservative 61; Mismatches 211; Indels 113; Gaps 20;
QY 36 INSTGPIGSRALFTPYR-----NSMADSGDNRASDVPGLPVPMR-LAASEITLND 85
Db 1464 insgkp-ghmaidqpvklliegkvnltvavtddsedntttnitvyyepkktlaaptvtps- 1521
QY 86 GFEVLHDHGPLDILNRQIGSSVFRVETQEDGKHIAVQQRNGVTSVVLSDQEVARLOSID 145
Db 1522 -----ttepaqvtvltanaaa-----tgetvqysadgkkyqdv-----aagvtit 1563
QY 146 PEGKDFVFTGGGGAGHAMVTVASDI-----TEARQRIELLEPKGTGSKGAGES 197
Db 1564 angtfkfstdlygnespavyvvttnikaddpaqlqaakqelnlliasaktlasakkydd 1623
QY 198 KGVGELRESNSGAENTTETSTSTSLR-SDPKLMTALGTATGLTGLAATGIVQALAL 256
Db 1624 atttalaatcqkaq-taltdqtnasvdsitganrdlqtainglaaklpadkktallnqlqs 1682
QY 257 TPE---PDSPTTTDPDAASATETATRDQLTKEA---FQNPDMQKNVIDELGNAPISGVL 310
Db 1683 vkdaigtldlgnqtdp--stgkttfaalddivagaqagtqtdqqlqatlak-----i 1731
QY 311 KDDVVAETEQAKAAGE-----EAKQQAII-----ENNAQAQKDYDEQQAQR 351
Db 1732 ldevlaklaegikaatpaevgnakdaatgktwyadiadtltsgqasadasdklahlqalq 1791
QY 352 QEELKVSS-----GAGYGLSGALILGGGIGVAVTA-----ALHRKNQP 389
Db 1792 slkckvaavaeaaaktvgkgdgtgtsdkgggqgtpapapadtgdkdgdsgqpsgggnip 1851
QY 390 VEQTTTTTTTTTSAR-----TVENKPNANTPAQGNVD---TPGSEDTMESRRSSM- 438
Db 1852 tkpattstttdtdtrngqitsgtsdkgggqgtpapapgdigdkdgdsgqpsgggnip 1911
QY 439 ----ASTSSFFDFTSSIGG 453
Db 1912 tnpattstttdtdtdtrng 1930

Search completed: September 27, 2001, 14:22:28
Job time: 124 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:21:04 ; Search time 21.2 seconds
(without alignments)
542.925 Million cell updates/sec

Title: US-09-189-415A-4

Perfect score: 2851

Sequence: 1 MP1GNLGHNNVNSIPPAP.....SNSAVNTSNNPPAGSHREV 559

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCRTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 125.5 | 4.4 | 461 | 1 | US-08-186-222-2 |
| 2 | 125 | 4.4 | 1702 | 4 | US-08-296-791-5 |
| 3 | 125 | 4.4 | 1702 | 5 | PCT-US95-10661A-5 |
| 4 | 124.5 | 4.4 | 1529 | 2 | US-08-728-470-10 |
| 5 | 124.5 | 4.4 | 1529 | 4 | US-08-719-641-10 |
| 6 | 124.5 | 4.4 | 1600 | 2 | US-08-617-697-10 |
| 7 | 123 | 4.3 | 658 | 1 | US-08-409-995-5 |
| 8 | 123 | 4.3 | 658 | 3 | US-08-685-467-5 |
| 9 | 123 | 4.3 | 658 | 4 | US-08-913-942-5 |
| 10 | 123 | 4.3 | 1098 | 1 | US-08-409-995-2 |
| 11 | 123 | 4.3 | 1098 | 3 | US-08-685-467-2 |
| 12 | 123 | 4.3 | 1098 | 4 | US-09-377-155-32 |
| 13 | 123 | 4.3 | 1098 | 4 | US-08-913-942-2 |
| 14 | 123 | 4.3 | 1848 | 4 | US-08-296-791-6 |
| 15 | 123 | 4.3 | 1848 | 5 | PCT-US95-10661A-6 |
| 16 | 123 | 4.3 | 2035 | 1 | US-08-046-585-5 |
| 17 | 123 | 4.3 | 2035 | 1 | US-08-393-703-5 |
| 18 | 123 | 4.3 | 2035 | 5 | PCT-US93-11721-5 |
| 19 | 120 | 4.2 | 675 | 1 | US-08-386-495-10 |
| 20 | 120 | 4.2 | 675 | 5 | PCT-US96-02331-10 |
| 21 | 120 | 4.2 | 842 | 5 | PCT-US96-02331-15 |
| 22 | 119 | 4.2 | 1536 | 1 | US-08-038-682-2 |
| 23 | 119 | 4.2 | 1536 | 1 | US-08-302-832-2 |
| 24 | 119 | 4.2 | 1536 | 2 | US-08-530-198-2 |
| 25 | 119 | 4.2 | 1536 | 2 | US-08-469-880-2 |
| 26 | 119 | 4.2 | 1536 | 2 | US-08-728-470-2 |
| 27 | 119 | 4.2 | 1536 | 2 | US-08-617-697-2 |

| | | | | | | |
|----|-------|-----|------|---|-------------------|--------------------|
| 28 | 119 | 4.2 | 1536 | 4 | US-08-719-641-2 | Sequence 2, Appli |
| 29 | 119 | 4.2 | 1581 | 4 | US-09-110-517-2 | Sequence 2, Appli |
| 30 | 118 | 4.1 | 1780 | 1 | US-08-769-309A-5 | Sequence 5, Appli |
| 31 | 118 | 4.1 | 1780 | 3 | US-08-994-570-5 | Sequence 5, Appli |
| 32 | 117.5 | 4.1 | 694 | 3 | US-08-559-397A-31 | Sequence 31, Appli |
| 33 | 117.5 | 4.1 | 1026 | 2 | US-08-614-377A-7 | Sequence 7, Appli |
| 34 | 117.5 | 4.1 | 1026 | 4 | US-09-142-648B-7 | Sequence 7, Appli |
| 35 | 116 | 4.1 | 546 | 3 | US-08-935-855-20 | Sequence 20, Appli |
| 36 | 115.5 | 4.1 | 985 | 5 | PCT-US96-03916-6 | Sequence 6, Appli |
| 37 | 115.5 | 4.1 | 985 | 5 | PCT-US96-03916-66 | Sequence 66, Appli |
| 38 | 115.5 | 4.1 | 1338 | 2 | US-08-728-470-9 | Sequence 9, Appli |
| 39 | 115.5 | 4.1 | 1338 | 4 | US-08-719-641-9 | Sequence 9, Appli |
| 40 | 115.5 | 4.1 | 1599 | 2 | US-08-617-697-9 | Sequence 9, Appli |
| 41 | 115 | 4.0 | 484 | 1 | US-08-127-499A-26 | Sequence 26, Appli |
| 42 | 115 | 4.0 | 484 | 1 | US-08-482-847-26 | Sequence 26, Appli |
| 43 | 111.5 | 3.9 | 631 | 4 | US-08-847-065-25 | Sequence 25, Appli |
| 44 | 111.5 | 3.9 | 1026 | 1 | US-08-194-290-7 | Sequence 7, Appli |
| 45 | 111 | 3.9 | 2353 | 4 | US-09-377-155-33 | Sequence 33, Appli |

ALIGNMENTS

RESULT 1
US-08-186-222-2
; Sequence 2, Application US/08186222
; Patent No. 5559007
; GENERAL INFORMATION:
; APPLICANT: Suri, Bruno
; TITLE OF INVENTION: Bacterial Vectors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/186,222
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,205
; FILING DATE: 19-MAR-1991
; APPLICATION NUMBER: GB 9006400.7
; FILING DATE: 22-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Villamizar, JoAnn
; REGISTRATION NUMBER: 30,598
; REFERENCE/DOCKET NUMBER: 4-17994/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914)785-7121
; TELEFAX: (914)347-5769
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-186-222-2

Query Match 4.4%; Score 125.5; DB 1; Length 461;
Best Local Similarity 19.0%; Pred. No. 0.0051;
Matches 84; Conservative 70; Mismatches 160; Indels 127; Gaps 16;

REFERENCE/DOCKET NUMBER: FP-59941/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1702 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-5

Query Match 4.4%; Score 125; DB 5; Length 1702;
Best Local Similarity 17.8%; Pred. No. 0.044;
Matches 87; Conservative 62; Mismatches 180; Indels 160; Gaps 17;
QY 10 PNVNNSI-----PPAPPLPSQTDGAGRGQLNSTGPLGSRALFTPVNRSMADSGDN 61
DB 1017 PSNNEELARVETVPVPPAPATSETTE-----TVAENSKQES--- 1053
QY 62 RASDVPGLVPMRLAASITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAV 121
DB 1054 -----KTVERNEQDATETTAQNG-EVAEAKPSVKANTQTN-----EVAQSGSETEE 1099
QY 122 GQRNGVETSVLSDQYARLQSIDPEGRDKFVFTGRCGAGHAMVTVASDITEARQRILE 181
DB 1100 TQTTEIKETAKVEKEEKAKEKAVEK-----DEIQEAPQMASE 1141
QY 182 LLEPKGTGSKGAGESKGVGELR-ESNSGAENTT-----ETQTSSTSSL 225
DB 1142 TSPKQAKPAKEVSTDTKVEETQVQAQPTQSTYVAAEATSPNSKPABETOPSEKTN-- 1199
QY 226 RSDPKLWALGVATGLIGLANTGIVQALALPEPSPPTTDPDAAASATETATRDQLTK 285
DB 1200 -AEP-----VTPVVKQNTENTTDQPTEREKTAKVETKQEPPOVAS 1241
QY 286 EAFQPNQKV-----NIDELGNAIPSGVLKDDVVANIEEQAKAAGEAQOATE 335
DB 1242 QASPKQEQSETVQPAVLSESNVPTVNN-----EEVQAQLOTQTSAT-VSTKQPAPE 1293
QY 336 NNAQAQKYYDEQAQRQELKVSAGYGLSGALILGGGIGVAVTAALHRKNQPVQETTT 395
DB 1294 NS-----INTGSATAITAEKSKDPQTETAA 1320
QY 396 TTTTTSSTARSV-ENKANNTPA-----QGNVDTPGSEDTMESRSSMASTSTFPD 447
DB 1321 STEDASQHKANTVADNSVANNSESSEPKSRRRRSISQP--QETSABETTAASDTETIAD 1378
QY 448 TSSIGGPCR 456
DB 1379 NSKRKPNR 1387

RESULT 4
US-08-728-470-10
Sequence 10, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

Query Match 4.4%; Score 124.5; DB 2; Length 1529;
Best Local Similarity 20.2%; Pred. No. 0.041;
Matches 123; Conservative 76; Mismatches 233; Indels 177; Gaps 28;

QY 26 TDGAGS-----RGOLINSTGPLGSRALFTPVNRSMADS-----GDNRASDVPGLPVPNMR 75
DB 882 TTNASGTQKTIINGNITNEKGLDN-----TKNIKADAEIQIGNISOKEGNLTISSOK 934
QY 76 L-AASEITLNDGFE-VLHDHGPLDTLNRQIGSSVFRVE-----TQEDGKHIA 120
DB 935 VNITQITIKAGVEGSRSDSEANANLTIQTKELAGDLNISGFENKAEITAKNGSGLT 994
QY 121 VQRNGVETSVLSDQYARLQSIDPEGRDKFVFTGRCGAGHAMVTVASDITEARQRIL 180
DB 995 IGNASGNAD-----AKVTFD-KVKDSKISTDGHN-----VTLSNV----- 1031
QY 181 ELLEPKGTGSKGAGESKGVGELRESNSGAENTTETQTSSTSSLSRSDPKLWALGTAVT 240
DB 1032 -----KTSNGSSNAGNDNSTG-----LTISAKDVTNNVNTSHKTI-----NISAAAGNVTT 1078
QY 241 --GLIGLAATGIVQALALPEPSPPTTDPDAAASATETATRDQLTK-AFON----- 290
DB 1079 KEGTINATGSEVTA-----QNGTIKGNITSQNVTVTATENLVTTENAVINATSGTVN 1133
QY 291 -----PDNQKVNIDELGNAIP-SGVLKDDVVANIEEQAKAAGEAQOAIEN 336
DB 1134 ISTKTDIKGGIESTSGNVTNATSGNTLKVSNITGQDVTVTADAGALTTTAGSTISATTG 1193
QY 337 NAAQAQKYYDEQAQRQELKVSAGYGLSGALILGGGIG---VAVTAALHRKNQPVQEQ 392
DB 1194 NANITTKTGDINGK-----VESSSGSVTLVATGATLAVGNISGNVTVITADSGKLTSTVGS 1249
QY 393 TTTTTSSTARSV-ENKANNTPAQGNVDTPGSEDTMESRSSMASTSTFPD----- 435
DB 1250 TINGTNSVTTSSQSGDIECTISGNTVNTASTGDL-TIGNSAKVKAENGAATLTAESGKL 1308
QY 436 -----SSMASTSTSTFTSSIGGPPRIRMLKHKRMRRCRLLILIRLFRWTGIG 486
DB 1309 TQTGSSITSSNGQTTLTAKD-SSIAG-----N 1335
QY 487 ISVYSTIQHPPTDNGARLLGNPSAGIQSTYARLALSGGLRHDGGLTGSGNSAVNT 546

Db 1336 INAAVTLNTGLTTTGDSKI--NATSGTLTINAKDA-----KLDGAASGDRVTVNA 1386
QY 547 SNNPPAPGS 555
Db 1387 TN---ASGS 1392
RESULT 5
US-08-719-641-10
; Sequence 10, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-719-641-10
Query Match 4.4%; Score 124.5; DB 4; Length 1529;
Best Local Similarity 20.2%; Pred. No. 0.041;
Matches 123; Conservative 76; Mismatches 233; Indels 177; Gaps 28;
QY 26 TDGAGG-----RGQLINSTGPGLSRALETPVRNSMADS-----GDNRASDVFPCLPVNPMR 75
Db 882 TTWASGTQKTIINGNITNEKGLN-----IKNKADAEIOIGGNISQKEGNLTISDDK 934
QY 76 L-AASETLNDGFE-VLHDGHPDLDLNRQIGSVFVFE-----TQEDGKHIA 120
Db 935 VNITNQITKAGVGGRSDSSEAEANLTIQTKELAGLDNLSGPNKAEITAKNGSDLT 994
QY 121 VGORNGVETSVLSDQEYARLQSIDPEGKDFVFTGGRGAGHAMVTVASDITEARQRL 180

Db 995 IGNASGNAD-----AKVTFD-KVKDSKISTDGHN-----VTLNSEV----- 1031
QY 181 ELLEPKGTGESKGAGSKGVGELRESNSGAENTETQTSTSTSSLRSDPKMLALGTVAT 240
Db 1032 -----KTSNGSSNAGNDNSTG-----LTISAKDVTNNNVTSHKTI-----NISAAAAGNVTT 1078
QY 241 --GLIGLAATGIVOALALTPEDSPPTTDPDAAASATETATRDOLTKE-AFON----- 290
Db 1079 KEGTTINATTGSVEVTA-----QNGTIKGNITSQNVTVTATENLVTTENAVINATSGTVN 1133
QY 291 -----PDNOKYNIDELGNAIP-SGVLKDDVVANIEEQAKAAGEEAKQQAEN 336
Db 1134 ISTKTGDIKGIESTSGNVNITAGSNITLKVNSITGODVTVTADAGALTTAGSTISATTG 1193
QY 337 NAAQAQKYDEQQAQROELKVVSSGA-GYGLSGALIIGGGIG-----VAVTAALHRKNOPVEQ 392
Db 1194 NANITTKTGIDNGK-----VSSSGSVTLVATGATLAVGNISGNTVTITADSGKLTSTVGS 1249
QY 393 TTTTNTTTTTSAR-----TVENKPNANTPAOGNVDTPGSEDTMESRR----- 435
Db 1250 TINGTNSVTTSQSGDIEGTISGNTVNTASTGDL-TIGNSAKVEAKNGAATLTASGKL 1308
QY 436 -----SSMASTSTFPDTSIGGPCRIIRMLKLRHRCRLLILIRLRIWGIQ 486
Db 1309 TTQTGSSITSSNGQTLTAKD-SSIAG-----N 1335
QY 487 ISVYVSTIQHPDRTDNGARLLGNPSAGIOSTYARLALSGGLRHDGGLTGGSNSAVNT 546
Db 1336 INAAVTLNTGLTTTGDSKI--NATSGTLTINAKDA-----KLDGAASGDRVTVNA 1386
QY 547 SNNPPAPGS 555
Db 1387 TN---ASGS 1392
RESULT 6
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-10

Query Match 4.4%; Score 124.5; DB 2; Length 1600;
Best Local Similarity 20.2%; Pred. No. 0.044;
Matches 123; Conservative 76; Mismatches 233; Indels 177; Gaps 28;

QY 26 TDGAGG-----RGLNSTGPLSRALFTVPRNSMADS-----GDNRASDPGLPVNPMR 75
DB 952 TTNASGTQKTINGNITNEKDLN-----IKNIKADAEIQIGNISQKEGNTLTISSDK 1004
QY 76 L-AASEITLNDGFE-VLHDHGPLOTNLROIGSSVFRVE-----TOEDGKHIA 120
DB 1005 VNINQIIRKAGVEGRSDSSEAEANANLTIQTKELKLAGDLNISGFNKAETAKNGSLT 1064
QY 121 VQORNGVETSVVLSQDEYARLQSIDPEGKDFVFTGGRGGAGHAMVTVASDITEARQRIL 180
DB 1065 IGNASGGNAD-----AKVTFD-KVKDSKISTDGHN-----VTLNSEV----- 1101
QY 181 ELLEPKGTGKSGKAGESKVGELRESNGAENTTQTSTSSLSRSDPKLWALGTVAT 240
DB 1102 -----KTSNGSSNAGNDNSTG-----LTISAKDVTVNNVNTSHKTI-----NISAAGNVT 1148
QY 241 -GLIGLAATGVQALALTPEPDSPPTTDPDAASATETATRDQLTKE-AFQN----- 290
DB 1149 KEGTTINATGSEVTA-----QNGTIKGNITSQNVTVTATENLVTTENAVINATSGTVN 1203
QY 291 -----PDNOKVNIDELGNAP-SGVLKDDVYVANIEQAAKAAQAAIEN 336
DB 1204 LSTKTDIKGGIESISGNVITASGNTLKVNSITQDVTVTADAGALTTAGSTISATG 1263
QY 337 NAAQAKKYDEQAKRQEBLKVSSGA-GYGLSGALILGGIG-----YAVTAALHRKNQVPEQ 392
DB 1264 NANITTKTGDINGK-----VSSSGSVTLVATCATLAVGNISGNTVTITADSGKLTSTVGS 1319
QY 393 TTTTITTTTTTSAR-----TVENKPNANTPAQGNVDTPCSEDTEMSRR----- 435
DB 1320 TINGTNSVTTSSQSDIGETISGNTVNTVNTAGDGL-TIGNAKVEAKNGAATLAESEGL 1378
QY 436 -----SSMASTSTFFDPTSSIGGCPRIIRMLKHKRMIRRCRLILIRLFRINGIQ 486
DB 1379 TTQTGSSITSSNGQTTAKD-SSIAG-----N 1405
QY 487 ISVYSTTQHPPRDTTNGARLLGNPSAGIOSTYARLALSGLRHDMGGLTGGNSAVNT 546
DB 1406 INAAVNTLTGTLTTGDSKI--NATSGTLTINAKDA-----KLDGAASGDRTVVNA 1456
QY 547 SNNPAPGS 555
DB 1457 TN---ASGS 1462

RESULT 7
US-08-409-995-5
; Sequence 5, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA

COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 658 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-5

Query Match 4.3%; Score 123; DB 1; Length 658;
Best Local Similarity 22.4%; Pred. No. 0.015;
Matches 111; Conservative 58; Mismatches 171; Indels 156; Gaps 25;

QY 10 PNVNNSIPPAPLPESQTDG-----ACGRQLNSTGPLSRALFTVPRN--SWADSGD 60
DB 194 PKVN-----VTSTDGLKFAQDAAGANGDTTVHLNGIGSTLTDTLVGSPATHIDGGD 245
QY 61 -----NRASDV-----PGLPVNPMRLAASEITLNDGFEVLHDHGPDLTLNRQIGSSVFRV 110
DB 246 QSTHYTRAASIKDVLNAGWNKGVK-AGSTTGQSENDFVHYTYDFELSDATETTTTV 304
QY 111 ETQEDGKHIAVQORNGVETSVVLSQDEYARLQSIDPEGKDFVFTGGRGGAGHAMVTVAS 170
DB 305 DSKENGKRTEV--KIGAKTSVI-----KEKDGKLTG----- 334
QY 171 DITEARQRILELLEPKGTGKSGKAGESKVGELRESNGAENTTQTSTSSLSRSDPK 230
DB 335 ---RANK---ETNKVDGANATEDADEGKLVAKDVIDAVNKTGWRIKTTDANGONGD-- 386
QY 231 LWLALGTVATGLIGLAATGIVQALALTPEPDSPPT-----TOPDAAASATETAT 279
DB 387 ----PATVASGNTVTFASGNGTTATVNGTGTITVKYDAKVDGDKLGDGDKTAADT-TA- 440
QY 280 RDQLTKEAFQNPNDOKVNIDELGNALPSGVLKDDVYVANIEQ-----AK----- 323
DB 441 ---LTVNDGNKANN-----PRGKVAD--VASTDEKKLVAKGLVTALNLSWT 483
QY 324 AAGEBAKQQAIENTNAQAQ--KKYDEQAKRQEBLKV-SSGAG--YGLSGALILGGIGVA 378
DB 484 TTAEEADGGTLDGNASEQEVKAGDKVTFKAGKNLKVKEGANFTYSLQDALT---GL--- 537
QY 379 VTAALHRKNQPVQETTTTTTTTTTTSARTVENKPA--NNTPAQGNVDTPCSEDTEMSRRSS 437
DB 538 -----TSITLGTGNGNKAETINKDGLTITPANG-----ACAN 569
QY 438 MASTSSTFFDTSISGG 453
DB 570 NANTISVTKDGISAGG 585

RESULT 8
US-08-685-467-5
; Sequence 5, Application US/08685467
; Patent No. 6060059

GENERAL INFORMATION:
APPLICANT: St. Gene III, Joseph W.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,467
FILING DATE: 22-JUL-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 658 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-685-467-5

Query Match 4.3%; Score 123; DB 3; Length 658;
Best Local Similarity 22.4%; Pred. No. 0.015;
Matches 111; Conservative 58; Mismatches 171; Indels 156; Gaps 25;

QY 10 PWNNSIPPAPPLPSQTDG-----AGRGQLNSTGPGLSRALFTPVNR--SMADSGD 60
DB 194 PKVN-----VTSTDTGLKFAKDAAGANGDTTVHLNGIGSTLTDTLVGSPTHIDGGD 245

QY 61 -----NRASDV-----PGLPVNPMRLAASEITLNDGFVLDHGHPLDTLNRQIGSSVFRV 110
DB 246 QSTHTYTRAAIKDVLNAGWNKGVK-AGSTTGQSENVDVHTYDVEFLSADTETTTTV 304

QY 111 ETQEDGKHIAVGQNGVTSVVLSDQEVARLQSIDPEGKDFVFTGGRGGAGHAMVTVAS 170
DB 305 DSKNGKRTEV--KIGAKTSVI-----KERDGLFTG----- 334

QY 171 DTEARQRIELLEPGTGESKAGSGVGLRESNSGAENTTETQTSTSTSLRSDPK 230
DB 335 ----KANK-----ETNKVDGANATEDGKGLVTAQKIDVAVNKTGWRIKVTDANGOND-- 386

QY 231 LWLALGTVAIGLIGLAATGIVQALATPPDPSFTT-----TDPDAASATETAT 279
DB 387 ----FATVAGTNVTFASNGTATVNTNGDTGITVKYDAKVGDLKLDGDKLAADT-TA- 440

QY 280 RDLTKFAFONPDNOKNIDELGNALIPSVLKDQDVANTEEQ-----AK----- 323
DB 441 ---LTVDNGKNANN-----PKGVAD--VASTDEKKLVTAQKGLVLTALNSLSWT 483

QY 324 AGEERAKQQAIEENNAQ--KKYDEQAKRQELKV--SSGAG--YGLSALLILGGGIGVA 378
DB 484 TTAEEADGGTLTDLGNASEQVEKAGDKVTFRAGNKLKVKQEGANFTYSLQDALI---GL--- 537

QY 379 VTAALHRKNQPVQETTTTTTTTTTTTTTTSARTVENKPA-NNTPAQGNVDTPGSEDTMESRSS 437
DB 538 -----TSITLGTGNGNGAKTEINKDGLTITPANG-----AGAN 569

QY 438 MASTSTSTFFDISSIGG 453
DB 570 NANTISVTKDGISAGG 585

RESULT 9
US-08-913-942-5
; Sequence 5, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 658 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-913-942-5

Query Match 4.3%; Score 123; DB 4; Length 658;
Best Local Similarity 22.4%; Pred. No. 0.015;
Matches 111; Conservative 58; Mismatches 171; Indels 156; Gaps 25;

QY 10 PWNNSIPPAPPLPSQTDG-----AGRGQLNSTGPGLSRALFTPVNR--SMADSGD 60
DB 194 PKVN-----VTSTDTGLKFAKDAAGANGDTTVHLNGIGSTLTDTLVGSPTHIDGGD 245

QY 61 -----NRASDV-----PGLPVNPMRLAASEITLNDGFVLDHGHPLDTLNRQIGSSVFRV 110
DB 246 QSTHTYTRAAIKDVLNAGWNKGVK-AGSTTGQSENVDVHTYDVEFLSADTETTTTV 304

QY 111 ETQEDGKHIAVGQNGVTSVVLSDQEVARLQSIDPEGKDFVFTGGRGGAGHAMVTVAS 170
DB 484 TTAEEADGGTLTDLGNASEQVEKAGDKVTFRAGNKLKVKQEGANFTYSLQDALI---GL--- 537

Db 305 DSKENGRKTEV--KIGAKTSVI-----KEKDGKLTG----- 334
 Qy 171 DITEARQILELLEPKGTGSGKAGESKGVGELRESNSGAENTTETQSTSTSLRSDPK 230
 Db 335 ---RANK---ETNKVDGANATEDAGKGLVAKDVIDAVNKGTGWRKTTDANGONGD-- 386
 Qy 231 LWALGTVATGLIGLAATGIQVQALALALTPEDSPPT-----TPDAAAASATETAT 279
 Db 387 ---FATVASGTNTVFASGNGTTATVTNGTIGTVKYDAKVGDLGLDGDKIAADT-TA- 440
 Qy 280 RDQLTKEAFQNDKQVNIDELGNALPSGVLKDDVVAIEEQ-----AK----- 323
 Db 441 ---LTVNDGKNANN-----PKGVAD--VASTDEKLVAKGLVTALNSLSWT 483
 Qy 324 AAGEAKQQAIAENNAQAQ--KKYDEQQAQROBELKV--SSGAG--YGLSGALILGGIGVA 378
 Db 484 TTAADGGLDGNASEQEVKAGDKVTFKAGKNLKVQEGANFTYSLQDALT---GL--- 537
 Qy 379 VTAALHRKNQVEQTTTTTTTTTTTTSARTVENKPA--NNTPAQGNVDTPGSEDPTMESRRSS 437
 Db 538 -----TSITLGTGNGNKAETKNGDGLTITPANG-----AGAN 569
 Qy 438 MASTSSTFFDTSSTGG 453
 Db 570 NANTISVTKDGISAGG 585

RESULT 10
 US-08-409-995-2
 ; Sequence 2, Application US/08409995
 ; Patent No. 5646259
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J.
 ; APPLICANT: St. Geme III, Joseph W.
 ; TITLE OF INVENTION: Hemophilus Adhesion Proteins
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 ; STREET: Four Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/409,995
 ; FILING DATE: 24-MAR-1995
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Silva, Robin M.
 ; REGISTRATION NUMBER: 38,304
 ; REFERENCE/DOCKET NUMBER: A-61053/RFT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEX: (415) 398-3249
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1098 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: unknown
 ; US-08-409-995-2

Query Match 4.38; Score 123; DB 1; Length 1098;
 Best Local Similarity 22.4%; Pred. No. 0.033;
 Matches 111; Conservative 58; Mismatches 171; Indels 156; Gaps 25;

Qy 10 PNVNNSIPPAPPLPSQTDG-----AGRGQLINSTGLGSRALFTVRN--SMADSGD 60
 Db 194 PKVN-----VSTDTGLKFAKDAAGANGDTTVHLNGIGSTLTDTLVGSPATHIDGDD 245
 Qy 61 ---NRASDV---PGLPVNPMRLAASEITLNDGFVFLHGHGDLPLDNLNRQIGSSVPRV 110
 Db 246 QSTHYTRAASIKDVLNAGWNKGVK-AGSTTCQSENVDVHTYDTVEFLSADTETTTTV 304
 Qy 111 ETQEDGKHIAVGQRNGVETSVLVSDQEVARLQSIDPECKDKFVETGGRGGAGHAMVTYAS 170
 Db 305 DSKENGRKTEV--KIGAKTSVI-----KEKDGKLTG----- 334
 Qy 171 DITEARQILELLEPKGTGSGKAGESKGVGELRESNSGAENTTETQSTSTSLRSDPK 230
 Db 335 ---RANK---ETNKVDGANATEDAGKGLVAKDVIDAVNKGTGWRKTTDANGONGD-- 386
 Qy 231 LWALGTVATGLIGLAATGIQVQALALALTPEDSPPT-----TPDAAAASATETAT 279
 Db 387 ---FATVASGTNTVFASGNGTTATVTNGTIGTVKYDAKVGDLGLDGDKIAADT-TA- 440
 Qy 280 RDQLTKEAFQNDKQVNIDELGNALPSGVLKDDVVAIEEQ-----AK----- 323
 Db 441 ---LTVNDGKNANN-----PKGVAD--VASTDEKLVAKGLVTALNSLSWT 483
 Qy 324 AAGEAKQQAIAENNAQAQ--KKYDEQQAQROBELKV--SSGAG--YGLSGALILGGIGVA 378
 Db 484 TTAADGGLDGNASEQEVKAGDKVTFKAGKNLKVQEGANFTYSLQDALT---GL--- 537
 Qy 379 VTAALHRKNQVEQTTTTTTTTTTTTSARTVENKPA--NNTPAQGNVDTPGSEDPTMESRRSS 437
 Db 538 -----TSITLGTGNGNKAETKNGDGLTITPANG-----AGAN 569
 Qy 438 MASTSSTFFDTSSTGG 453
 Db 570 NANTISVTKDGISAGG 585

RESULT 11
 US-08-685-467-2
 ; Sequence 2, Application US/08685467
 ; Patent No. 6060059
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Geme III, Joseph W.
 ; APPLICANT: Barenkamp, Stephen J.
 ; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 ; STREET: Four Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/685,467
 ; FILING DATE: 22-JUL-1996
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/409,995
 ; FILING DATE: 24-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Silva, Robin M.
 ; REGISTRATION NUMBER: 38,304
 ; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEX: (415) 398-3249
 ; TELEX: 910 277299

```

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-685-467-2

Query Match 4.3%; Score 123; DB 3; Length 1098;
Best Local Similarity 22.4%; Pred. No. 0.033;
Matches 111; Conservative 58; Mismatches 171; Indels 156; Gaps 25;

QY 10 PNVNNSIPPAPPLPSOTDG-----AGGRGQLINSTGPLGSRALFTFVRN--SMADSGD 60
DB 194 PKVN-----VTSTDTGLKFAKDAAGANGDTTVHLNGIGSTLTDTLVGSPATHIDGGD 245
QY 61 -----NRASDV-----PGLPVPNPMRLAASEITLNDGFEVLHDHGPDLTLNRQIGSSVFRV 110
DB 246 QSTHYTRAASIKDVLNAGWNKGVK-AGSTTGQSENVDVFVHTYDTVEFLSADTETTTTV 304
QY 111 ETQEDGKHIAVGORNGVETSVLSDDQYARLQSIDPEGDKFVFTGGRGGAGHAMVTVAS 170
DB 305 DSKENGKRTVE--KIGAKTSVI-----KEKDGKLFVG-----334
QY 171 DITAEARQRIELLEPEKGTGESKGAGESKGVGELRESNGAENTTETOTSTSTSLSDPK 230
DB 335 ---KANK---ETNKVDGANATEDADEGKGLVTAKDVIDAVNKTGWRIKTTDANGQNGD-- 386
QY 231 LWALGTVATGLIGLAATGIVQALALTPEDPSPTT-----TDPDAASATETAT 279
DB 387 ---FATVASTGNTVFASGNGTTATVTNGTGTGIVKYDAKVGDGLDKDKIAADT-TA- 440
QY 280 RDLTKEAFQNDKQVNIIDELGNAIPSGVLKDDVVAIEEQ-----AK-----323
DB 441 ---LTVNDGKNANN-----PKGKVD--VASTDEKLVTAAGLVLTALNSLSWT 483
QY 324 AAGEEAKQQAIAENNAQAQ--KKYDEQQAQKQOEELKV--SSGAG--YGLSGALILGGGIGVA 378
DB 484 TTAEEADGGTLDGNASEQEVKAGDKVTFKAGKNLKVQEGANFTYSLQDALT---GL---537
QY 379 VTAALHRKNQPVQETTTTTTTTTTTTTSARTVENKPA-NNTPAQGNVDTPGSEDTEMESRRSS 437
DB 538 -----TSITLGTGNGNGAKTEINKDGLTITPANG-----AGAN 569
QY 438 MASTSTSTFTDSSIGG 453
DB 570 NANTISVTKDGISAGG 585

RESULT 13
US-09-377-155-32
; Sequence 2, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: PCT/US96/4031
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296.791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1098 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-913-942-2

Query Match 4.3%; Score 123; DB 4; Length 1098;
Best Local Similarity 22.4%; Pred. No. 0.033;
Matches 111; Conservative 58; Mismatches 171; Indels 156; Gaps 25;
Qy 10 PNVNNSIPPAPPLPSQTDG-----AGRGOLINSTGLSGRALFTVPVN--SMADSGD 60
Db 194 PKVN-----VSTTDGLKFAKDAAGANGDTTVHLNGIGSTLTLTLVGSPTHIDGDD 245
Qy 61 -----NRADV-----PGLVPNPMRLAASEITLNDGFVHLHDGDLTLNRQIGSSVFRV 110
Db 246 QSTHYTRAASIKDVLNAGWNKGVK-AGSTTGQSENVDFVHTYDVFELSADTETTTTVV 304
Qy 111 ETQDGHKHAIVGQNGVTSVLSDOEYARLOSIDPEGKDFVFTGGRGAGHAMVTVAS 170
Db 305 DSKNGKRTVE--KIGAKTSVI-----KEKDGKLEFG----- 334
Qy 171 DITEARQILELLEPKGTGESKAGESKVGELRESNGAENTTETQSTSTSLRSDPK 230
Db 335 ---KANK---ETNKVDGANATEDAGEGLTKAKDVIDAVNKTGHRKITTANGONGD-- 386
Qy 231 LWALGTVATGLIGLAATGIVQALALTPEPSPPT-----TPDAAAASATETAT 279
Db 387 ---FATVASGTNTVFASGNGTTATVTNGTDGTTVKYDAKVGDLKLDGDKIAADT-TA- 440
Qy 280 RDQLTKEAFQNPDKQKVIDELGNALPSGLKDDVVANIEEQ-----AK----- 323
Db 441 ---LTVNDGKNANN-----PKGVAD---VASTDEKLVAKGLVYALNLSWT 483
Qy 324 AAGERAKQQAIEENNAQAQ--KKYDQQAQKROEELKV-SSGAG--YGLSGALILGGGIGVA 378
Db 484 TTAADAGGTLDGNASEQEVKAGDKVTKAGKNLKVKGAGNFYSLQDALT---GL--- 537
Qy 379 VTAALHRKNQVEQTTTTTTTTTTTTSARTVENKKA-NNTPAGNVDPFGSEDTMESRRSS 437
Db 538 -----TSITLGTGNGNAKTEINKDGLTITPANG-----AGAN 569
Qy 438 MASTSSSTFFDSSIGG 453
Db 570 NANTISVTKDGISAGG 585

RESULT 14
US-08-296-791-6
Sequence 6, Application US/08296791
Patent No. 6245337
GENERAL INFORMATION:
APPLICANT: St. Gene III, Joseph W.
APPLICANT: Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert

STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296.791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-296-791-6

Query Match 4.3%; Score 123; DB 4; Length 1848;
Best Local Similarity 21.3%; Pred. No. 0.076;
Matches 103; Conservative 52; Mismatches 174; Indels 154; Gaps 23;
Qy 17 PPAP-----PLPSQTDGAGRGOLINSTGLSGRALFTVPVRNSMA 56
Db 1038 PPAPATAISAEQETPRPAETAQAQAMEETNTANS-----TETAPKSDTATQENPNS 1092
Qy 57 DSGDNRAADVGLPNPMLA-----ASEITLNDGFVHLHDGDLTLNRQIGSSVFRVE 111
Db 1093 VPSETTEKVAENPPQENETVAKNEQEAETPTQNG-EVAKEDQPTVEANTQTNEA---- 1146
Qy 112 TQEDGKHIAVGQNGVETSVLSDOEYARLOSIDPEGKDFVFTGGRGAGHAMVTVASD 171
Db 1147 TQSEK-----TEETQTAETKSEPTES-----VTVSEN 1174
Qy 172 ITEARQILELLEPKGTGESKAGESKVGELRESNGAENTTETQSTSTSSLSRDPKL 231
Db 1175 -----QPEKTVSQ--STEDKVVVE-KEEKAKVE-TEETQKAPQVTS-KEPKK- 1216
Qy 232 WLALGTVATGLIGLAATGIVQALALTPEPDS-PTTTPDPA-----ASATETATR 280
Db 1217 -----QA-----EPAPEEVPDTNAAEEAQAALQOQPTTVAEAETTS 1253
Qy 281 DQLTKEAFQNPDKQKVIDELGNALPSGLKDDVVANIEEQAKAAGEAKQQAIEENNAQA 340
Db 1254 NSKPAEETQOP-SEKTNAPVTPPVSENTATQP--TETEETAKVE-KEKTVQVPOVASQE 1309
Qy 341 OKKYDEQQAQKROEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKN-----QPVEQT 396
Db 1310 SPKQEQPAKPAQATKPAEP-----ARENVLTKNVGEPQPAQPOTQ 1353
Qy 397 TTTTTSARTVENKKA-----NNTPAGNVDPFGSEDTMESRRSSMASTSTFFD 447
Db 1354 STAVPTTGTAAANSKPAKPAQAKPQEPARENVSTVNTKEP-QSOTSATVSTEQPAKE 1412
Qy 448 TSS 450
Db 1413 TSS 1415
RESULT 15
PCT-US95-10661A-6

```
Sequence 6, Application PC/TUS9510661A
GENERAL INFORMATION:
APPLICANT: Washington University, et al.
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 16-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: FP-59941/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-6
```

```
Query Match 4.3%; Score 123; DB 5; Length 1848;
Best Local Similarity 21.3%; Pred. No. 0.076;
Matches 103; Conservative 52; Mismatches 174; Indels 154; Gaps 23;

QY 17 PPAP-----PLPSQTDGAGRGGLINSTGFLGSRALFTPVNRNSMA 56
DB 1038 PPAPATESAIASEQPETRAPETAQPAEETNTANS-----TETAPKSDTATQTENPNSES 1092,
QY 57 DSGDNRASDVPLGPNMRLA-----ASEITLNDGFEVLHDHGPDLTLNRIGSSVFRVE 111
DB 1093 VPSEETKVAENPPQENETVAKNEQEATEPTPQNG-EVAKEDQPTVEANTQTNEA-----1146
QY 112 TQEDGKHIAVGORNGVTSVLSDOEYARLASIDPEGKDKFVFTGGRGAGHAMVTVASD 171
DB 1147 TQSEKG-----TEETQTATKSEPTES-----VTVSEN 1174
QY 172 ITEARQIULELEPKGTGSKGAGSGVGLRESNGAENTTETOTSTSSLSRSDPKL 231
DB 1175 -----QPEKTVSQ--STEDKVVVE-KEEKAKVE-TEETQKAPQVTS-KEPKK- 1216
QY 232 WLALGTAVATGLIGLAATGIVQALALTPEDS-PTTTDPDAA-----ASATETATR 280
DB 1217 -----QA---EPAPEEVPTDTNAEEAQAQQOQTPTTVAAAEETTSP 1253
QY 281 DOLTKEAFONPNQKNIDELGNALPSGVLDKDDVVANTEEQAAGEAKQAQAIENNAQA 340
DB 1254 NKPAAEETQOP-SEKTNAPETVTPVSENATQOP--TETEETAKVE-KETQEVPPQVASOE 1309
QY 341 QKDYDQQAQRQOEELKLVSSGAGYGLSGALILGGIGGVAVTAALHRKN-----QPVEQTTTT 396
DB 1310 SPKQEQPAKPPQATKPPAEP-----ARENVLTTKNVGEPPQAPQPTQ 1353
```

```
Search completed: September 27, 2001, 14:21:08
Job time: 44 sec

QY 397 TTTTTSARTVENKPA-----NNTPAQGNVDTPGSEDMTESRRSSMASTSTSTFFD 447
DB 1354 STAVPTTGETAANSKPAKPAQAKPQTEPARENVTNTKEP-QSQTSATVSTEQPAKE 1412
QY 448 TSS 450
DB 1413 TSS 1415
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:21:42 ; Search time 25.62 Seconds
(without alignments)
1662.044 Million cell updates/sec

Title: US-09-189-415a-4
Perfect score: 2851
Sequence: 1 MPICNLGHNPNVNNISIPPAP.....SNSAVNTSNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68: *
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------------------------|
| 1 | 2624.5 | 92.1 | 558 | 2 E86045 | probable translocase |
| 2 | 170 | 6.0 | 1229 | 2 T25697 | hypothetical prote |
| 3 | 165.5 | 5.8 | 2232 | 2 T34434 | hypothetical prote |
| 4 | 162.5 | 5.7 | 1291 | 2 T13389 | hypothetical prote |
| 5 | 152.5 | 5.3 | 973 | 2 C85693 | probable membrane related to C2H2 zi |
| 6 | 148.5 | 5.2 | 770 | 2 T51024 | probable serine/th |
| 7 | 148 | 5.2 | 1192 | 2 T18611 | probable tail fibe |
| 8 | 146.5 | 5.1 | 1122 | 2 G54887 | hypothetical prote |
| 9 | 144.5 | 5.1 | 3507 | 2 T34513 | hypothetical prote |
| 10 | 143.5 | 5.0 | 1275 | 2 T33369 | hypothetical prote |
| 11 | 143.5 | 5.0 | 1630 | 2 A53577 | ascites sialoglyco |
| 12 | 142 | 5.0 | 563 | 2 A36054 | mucin homolog - bo |
| 13 | 141.5 | 5.0 | 1962 | 2 A32634 | lactocepin (EC 3.4 |
| 14 | 140.5 | 4.9 | 1026 | 1 A40315 | maternal effect pr |
| 15 | 140.5 | 4.9 | 1829 | 2 T24583 | hypothetical prote |
| 16 | 139 | 4.9 | 2468 | 2 A83412 | hypothetical prote |
| 17 | 137 | 4.8 | 837 | 2 T02761 | outer arm dynein i |
| 18 | 137 | 4.8 | 4558 | 2 C82199 | RTX toxin RTXA VC1 |
| 19 | 136 | 4.8 | 1005 | 2 H85611 | probable adhesin z |
| 20 | 136 | 4.8 | 1005 | 2 B85663 | probable adhesin z |
| 21 | 134 | 4.7 | 2514 | 2 F81045 | hemagglutinin/hemo |
| 22 | 133.5 | 4.7 | 796 | 2 T21460 | hypothetical prote |
| 23 | 133 | 4.7 | 439 | 2 E71497 | hypothetical prote |
| 24 | 133 | 4.7 | 1589 | 2 T42233 | submaxillary mucin |
| 25 | 132.5 | 4.6 | 461 | 2 JN0097 | secreted 45K prote |
| 26 | 132.5 | 4.6 | 755 | 2 H85661 | Cr456 hypothetical |
| 27 | 132.5 | 4.6 | 755 | 2 B72061 | hypothetical prote |
| 28 | 132.5 | 4.6 | 1258 | 2 JQ0188 | ice nucleation pro |
| 29 | 132.5 | 4.6 | 1459 | 2 T32271 | hypothetical prote |

30 131 4.6 690 2 T31597 hypothetical prote
31 131 4.6 1083 2 JC2300 cell surface glyco
32 131 4.6 2660 2 E85822 probable invasin z
33 130.5 4.6 829 2 D71485 probable adenylate
34 130.5 4.6 1034 2 JC2143 ice nucleation act
35 130 4.6 1063 2 D86731 hypothetical prote
36 130 4.6 1077 2 A44067 serine-rich protei
37 130 4.6 2055 2 T31110 extracellular matr
38 129.5 4.5 2059 2 D82671 surface protein XF
39 129.5 4.5 2411 2 A46299 tyrosine kinase su
40 129 4.5 666 2 G82111 flagellar hook-ass
41 129 4.5 784 2 S26638 SPR-1 protein - hu
42 129 4.5 1460 2 D1675 polymorphic membra
43 129 4.5 3190 2 T13828 CREB-binding prote
44 128.5 4.5 532 2 T06029 hypothetical prote
45 127.5 4.5 436 2 G81655 conserved hypothet

RESULT 1
E86045
probable translocated intimin receptor protein tir [imported] - Escherichia coli (str C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E86045
K:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: E86045
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-558 <STO>
A:Cross-references: GB:AB005174; NID:gl2518449; PIDN:AAG58825.1; GSPDB:GN00145; UWGP:
C:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: tir

Query Match 92.1%; Score 2624.5; DB 2; Length 558;
Best Local Similarity 94.0%; Pred. No. 3.6e-148;
Matches 528; Conservative 3; Mismatches 24; Indels 7; Gaps 3;

Qy 1 MPICNLGHNPNVNNISIPPAPLPSTQDGAGRGQLINSTGPGSLALFTPPVRNSMADSGD 60
Db 1 MPICNLGHNPNVNNISIPPAPLPSTQDGAGRGQLINSTGPGSLALFTPPVRNSMADSGD 60
Qy 61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPDLTLNRQIGSSVFRVETQEDGKHFA 120
Db 61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPDLTLNRQIGSSVFRVETQEDGKHFA 120
Qy 121 VQQRNGVETSVVLSQDEYARLQSIDPEGKDFVFTGGRGAGHAMVTVASDITEARQRL 180
Db 121 VQQRNGVETSVVLSQDEYARLQSIDPEGKDFVFTGGRGAGHAMVTVASDITEARQRL 180
Qy 181 ELLEPKGTGESKAGESKGVGELRESNGAENTTQTSTSTSSLRSDPKLMLALGTAV 240
Db 181 ELLEPKGTGESKAGESKGVGELRESNGAENTTQTSTSTSSLRSDPKLMLALGTAV 240
Qy 241 GLIGLAATGIVQALALTPEPDSPTTTDPDAASATETATRDQLTKEAFQNPQKNVIDE 300
Db 241 GLIGLAATGIVQALALTPEPDSPTTTDPDAASATETATRDQLTKEAFQNPQKNVIDE 300
Qy 301 LGNALPSGVKDDVVANTEEQAAGEAKQAIENNAQAOKKYDEQAQKROEELKVSSG 360
Db 301 LGNALPSGVKDDVVANTEEQAAGEAKQAIENNAQAOKKYDEQAQKROEELKVSSG 360
Qy 361 AGYGLSGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTTTSARTVENKPNNTPAQ 420
Db 361 AGYGLSGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTTTSARTVENKPNNTPAQ 420

ALIGNMENTS

Thu Sep 27 15:30:01 2001

QY 421 NVDTPGSEDTMESRRSSMASTSTFFDTSSIG---GPCRIRMLMLKHKRMIRRCRLILILI 477
Db 421 NVDTPGSEDTMESRRSSMASTSTFFDTSSIGVQNPYADVKTSL-HDSQVPTNSNTSV 479
QY 478 RLPRINGIQISVVYSTIOHPRTTONGARLLGNPSAGIOSTYARLALSGGLRHDHMGGLT 537
Db 480 QNM---GNTDSVVYSTIOHPRTTONGARLLGNPSAGIOSTYARLALSGGLRHDHMGGLT 536
QY 538 GGSNSAVNTSNPPAGSHRFV 559
Db 537 GGSNSAVNTSNPPAGSHRFV 558
RESULT 2
T25697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T25697
R:Fulton, B.
A:Description: The sequence of C. elegans cosmid F16F9.
A:Reference number: Z20071
A:Accession: T25697
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1229 <FUL>
A:Cross-references: EMBL:U67956; PIDN:AAB07691.1; GSPDB:GN00028; CESP:F16F9.2
A:Experimental source: strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP:F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3
Query Match 6.0%; Score 170; DB 2; Length 1229;
Best Local Similarity 22.8%; Pred. No. 0.025;
Matches 68; Conservative 44; Mismatches 122; Indels 64; Gaps 9;
QY 167 TVASDITEARQRIELLPKGTGSGKAGESKGVGELR-----ESNSGAENTETOTSTST 222
Db 597 SVSTKSTTKASTKEETPTTDEPTTSTSTTGKATPELSTSEETTTKLTITTEG 656
QY 223 SLRSDPKLWLALGVATGLIGLAATGIVQALALPEPDSPTTTPDPAASAATATRDQ 282
Db 657 STTTEP-----TTTAFAEASTGII---TTDEETSTSTTPELTSTKEIVTESA 704
QY 283 LRKEAF-----QNPDNQKVNDELGNALPSGVLKDDV-VANIEQAKAAGEAKQQ 332
Db 705 ITQTSVVSVESSTPQLPWRKAIWNKFKHNL--VLKKEKRLKKEKESTTSGSDS-- 760
QY 333 ATENNAQAQKKYDEQAKRQELKQVSSGAGYGLSGALILGGGIGVAVTAALHRKNQVPEQ 392
Db 761 --ETTTVAENTDEVTTTEKEV-----VQTPTTTEKSTTQEE 797
QY 393 TTTTITTTTTTTTGARTVENKPNANTPAQGNVDTPGSEDTMESRRSSMASTSTFFDTSS 450
Db 798 TTTTITTTTETKTSKTTTETPTTSES-----TTTITTSPTSTSTTDTSS 844
RESULT 3
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisel, C.; Gattung, S.
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-2232 <GEI>
A:Cross-references: EMBL:U0846; PIDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 20
Query Match 5.8%; Score 165.5; DB 2; Length 2232;
Best Local Similarity 19.3%; Pred. No. 0.1;
Matches 113; Conservative 81; Mismatches 230; Indels 161; Gaps 20;
QY 9 NPNVNSIPPAPLPQSQTGAGRGQLINSTGLPSRALFTPVNSM----- 55
Db 1343 SPSSISPVPTSSPIPTTFASSTSGSTISDVSVSTSL-APLSSSLPSTVPSSSTQSFSS 1401
QY 56 ABSGDNRASDVP-----GLPVNPMRLAASEITLNDGFVLDHGHGPDLTNLRIQIGSSVFR 109
Db 1402 TSEGSKASSSPVPSQTSSTPTNPTGSTESLLSSTISGSTQH---TTMSKASSGSTSP 1458
QY 110 VEQEDGKHIVG--QRNGVETSVLSDQYARLQSIDPEGDKKFVFTGGRGAGHA--- 164
Db 1459 STNSQTGSTVTMGSSSTSGVSTSSASTOPQMSSTSGSSAGSTVASSTASPAASSTAPSS 1518
QY 165 -----MVTVASDITEARQRIELLPKGTGSGKAGESKGVGELRESNGAENTTET 216
Db 1519 TGTMSSTSSGTGVTGTTSES-STTASASSQTGSTVTMGSSSTSGV-----STSSASTOP 1571
QY 217 QTSTSTSLRSDPKLWLALGVATGLIGLAATGIVQALALPEPDSPTTTPDPAASAAT 276
Db 1572 QMSTSQGS-----SAGSTVASSTAGLVTSTV-----PSSTGTMGSTSSGTGVS 1615
QY 277 TATRDQLTKFAQNPQKVNDELGNALPSGVLKDDV-VANIEQAKAAGEAKQQAIE 336
Db 1616 TISESSTTASA---SSQTGSTVTMGSSSTSGVSTSA----- 1649
QY 337 NAOAKKYDEQAKRQELKQVSSGAGYGLSGALILGGGIGVAVTAALHRKNQVPEOTTT 396
Db 1650 -----SSTQPMSTSQGS---AGSTVASSTGLVSTIV-----PSS 1684
QY 397 TTTTITTTTARTVENKPNANTPAQGNVDTPGSEDTMESRRSSMASTSTFFDTSSIGGPCR 456
Db 1685 TGTMGSTSSGTGVTGTTSESSTASASTGSTVTMGSSSTSGVSTSS-----ASSGQP-- 1737
QY 457 IRMLMLKHKRMIRRCRLILIRLFRINGIQIS-----VVYSTIOHPRTTONGA 506
Db 1738 -----QMSTSQGSAGSTVSVSSTASPAASSTAPSS 1768
QY 507 RLGNPSAG-IQSTYARLALSGGLRHDWGG-LTGGSNSAVNTSN 549
Db 1769 GTMSSTSGTGVSTMSQSSSTAASSTHTGTVTILGSSS---TSSN 1810
RESULT 4
T13389
hypothetical protein 115C2.10 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13389
R:Salles, C.; Valentti, P.; Darlamitsou, A.; Henderson, N.; Campbell, L.; Glover, D.
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17665
A:Accession: T13389
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1291 <CAT>
A:Cross-references: EMBL:AL031581; NID:e1320978; PID:e1426292; PIDN:CAA20894.2
C:Genetics:
A:Cross-references: FlyBase:FBgn0020381
A:Map position: X

A: Introns: 238/3; 1225/1
A: Note: EG:115C2.10

Query Match 5.7%; Score 162.5; DB 2; Length 1291;
Best Local Similarity 20.6%; Pred. No. 0.075;
Matches 99; Conservative 61; Mismatches 189; Indels 131; Gaps 18;

QY 5 NLGHNPN-VNNSIPPAPPLPSQTDG---AGGRCQLNSTGPGLSRALFTTPVRNSMADSGDN 61
DB 438 NSTSNSNTNDSTGPSETSSNGLVAGGAG-----GATGAAMLPTP---SQOSTGK 488
QY 62 RAS-----DVPGLPVNPMRLA-----ASEITLNDGFEVLHDHGPLDTLNQ 102
DB 489 EATAAVALLEKKLPNVVSPPLMKELRQKGTKYDAEMIMANAYQQQHH-----539
QY 103 IGSVFRVETQEDGKHIAVGRNGVETSVVLSQDYARLQSIDPEGKDFVFTGGRGAG 162
DB 540 --QHFFHHHHHHHHHGHQHAAGTAAVQMAAQKPG-----VGGTGAAG 588
QY 163 HAMVIVASDITEARQILLELLEPKGTGSGKAGSGVGLRESNSGAENTTETQSTST 222
DB 589 NAGATTVSSVA-----AGAGSEVNGRSTSLRKSrvns-----TSSI 627
QY 223 SSLSRDPKWLALGIVATGLIGLAATGIQALALATPEPDSPTTTPDAAASATETATRDQ 282
DB 628 STASADEVI-----APVVAASISLPKAPVVLMPCKPKPAOWATAALHQ 670
QY 283 LKPEAPQPNQKVNIDELGNAPSGVLKDDVYANTEEQAAGEEKQO-----A 333
DB 671 SQORQLRRSERQKELTD-----GESSD---TSSEQKK---EQKQDHLQPKMFS 716
QY 334 IENNAQAQKYDEQAQKROEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVQET 393
DB 717 LAEPQPEKS-EKQEQEQKRYTRNSAGRVGL-----VARLATAHNN-----TA 760
QY 394 TTTTITTTTTSARTV---ENKPNANTPAQGNVDTGSEDTMESRRSSMASTSTPFDTS 450
DB 761 TTNSSSSSKATTITNCNNHNSNRHNSLSRLSVKSRKPAPSEASIPSTSS 820
RESULT 5
C85693
probable membrane protein of prophage CP-933X Z1918 [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: C85693
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
killer, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; PMID:21074935; PMID:11206551
A:Accession: C85693
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-973 <STO>
A:Cross-references: GB:AE005174; NID:g12514847; PIDN:AAG56007.1; GSPDB:GN00145; UWGP:Z19
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1918

Query Match 5.3%; Score 152.5; DB 2; Length 973;
Best Local Similarity 22.1%; Pred. No. 0.2; Length 973;
Matches 87; Conservative 55; Mismatches 169; Indels 83; Gaps 14;

QY 79 SEITLNDGFEVLH-----DHGPLDTLNQISVFRVETQEDGKHIAVGO-----123
DB 61 SVILLVEGFPFPHAGTIVYEDSQP-GTNDLFLGAM-----TEDDVRPEALRFELWVEE 114
QY 124 --RNGVETSVVLSQDYARLQSIDPEGKDFVFTGGRGAGHAMVTVASDITEARQILE 181
DB 115 VARN---ASAVAQNTAAAKKSASDAS-----TSAREATHA--TDAADSARA-----156

QY 182 LLEPKGTGSGKAGSGVGLRESNSGAENTTETQSTSTSSLSRDPKWLALGTVATG 241
DB 157 -----ASTSAGQAASSAQSSASTASTKATEAKSAAAESEK-----SAAAT 201
QY 242 LIGLAATGIVQALALTPEDSPPTTTPDAAASATATATARDQLTKEAFQNPQKVNIDEL 301
DB 202 SAGAAKTSETNAVVOQSAAATSTASTATTKASEAASSARDASASKEAAKSETSAAS---257
QY 302 GNAIPSGVLKDDVYANTEEQAAGEEK-QQAIENNAQAQKYDEQAQKROEELKVSSG 360
DB 258 -----SASSAASSATAAGNSAKAAKTSETNAKSSSETAQAQSAAGSK--TA 303
QY 361 AGYGLSGALLGGIGVAVTAALHRKNQPVQETTTTTTTTTTTTTSARTVENKPNANTPAQ- 419
DB 304 AALSASAASTAGQASASATAA-----GKSAESAASASTATTKAGENTEQAASAAASSASA 359
QY 420 ---GNVDTGSEDTMESRRSSMASTSTPFDTS 450
DB 360 AKTSETNAKASETSAESSKTAASASSASSAS 393
RESULT 6
TS1024
related to C2H2 zinc finger transcription factor D-Spl [imported] - Neurospora crassa
N:Alternate names: protein B7F21.50
C:Species: Neurospora crassa
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: TS1024
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, July 2000
A:Reference number: Z25286
A:Accession: TS1024
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-770 <SCH>
A:Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.50
A:Experimental source: BAC clone B7F21; strain OR74A
C:Genetics:
A:Gene: NCSP:B7F21.50
A:Map position: 6
A:Introns: 117/1

Query Match 5.2%; Score 148.5; DB 2; Length 770;
Best Local Similarity 19.9%; Pred. No. 0.25;
Matches 109; Conservative 62; Mismatches 151; Indels 227; Gaps 25;

QY 5 NLGHNPN-VNNSIPPAPPLPSQTDGAGGRCQLNST---GPGLSR-----ALF 48
DB 219 NSGAIPKFLPSSVPKASAIKPKDKSDGSGG---LNGAQLGGIIGGATALLIIVIAAF 274
QY 49 TPVR--NSMADSGDNRAVDVGLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNQIGSS 106
DB 275 LIIRLKRIVESAMESKKGTSYHRSKASKTSQAQMEQSGRFLHVRAPSDTDNASADPL 334
QY 107 VERVETOEDGKHI-----AVGO--RNGVETSVVLS-----DQ 136
DB 335 MFMSETNTPGDHTTNASSLAGTPQCAHGVGDGSGIGRHSRSGSDTTTYMASHPGAGNNN 394
QY 137 EYARLQSIDP-----EGKD-----KFVFTGGRGAGHAMVTVASDITEARQILE 181
DB 395 NGSELASDPNRYFDGASPPPLPSSHTTSTTYAGCHNGHGM-----RESV--442
QY 182 LLEPKGTGSGKAGSGVGLRESNSGAENTTETQSTSTSSLSRDPKWLALGAT-----237
DB 443 -----DSQSTGLGYHYSTTRNQHQHHRNQSNASELSAD-----GSEITH 483
QY 238 -VATGLIG-----LAATGIV-----QALATPEPDS-----PTTTPD 269
DB 484 GVASPLVGGSSHARGASGTSRYTHHTSHSHSGVDVPELDSSGMFVELPATTTTTT 543
QY 270 AAASAT-----ETATRDQLTKEAFQNPDN-----QKVNIDELGNAIPSGVLKDDVVA 316

| | | | | | | |
|------------------------|--|---|--------------------------|-----------------|-------------|----------|
| QY | 146 | PEGKDKFVFTGGRGAGHAMVTVASDITAEARQILLELLEPKTGESKGAGSGKVGVELRE | 205 | | | |
| Db | 601 | -----GTRHGGVQMAQPT--SRQATISLLQPPSYKPPSNTTQIAQTPPLFN | 645 | | | |
| QY | 206 | SNSGANTTETQTSTSSLR--SDPKLWLALGT-----VATCLIG----- | 244 | | | |
| Db | 646 | RNSTA--TSSAAQPFSTGITGRKIADPKGRIPLNSTAVOGHRTATGAVANNGGIPSHRDH | 704 | | | |
| QY | 245 | -----LAATGIVQALALTPEPDSPTTTPDPAASATETATRDQLTKEAFO----- | 289 | | | |
| Db | 705 | AQOOQYMNQLTSTMSKLINKTPAAGGTAATSSSSSSSATSTA--PLOGSGSQISHAP | 761 | | | |
| QY | 290 | -----NPDNOKYNIDELGNAIPS-----GVLKDDVVANI-----EEQAKAAGEA | 329 | | | |
| Db | 762 | TEPVIREDDDENNSENQNGVPLIGVGQPTSPAVGQPTEDATSSSDKEQQOQKASSETP | 821 | | | |
| QY | 330 | KQAIENNAQAQKKYDEQQAQKROEFLKVSAGYGLSGALILGGGIGVAVTAALHRKNQP | 389 | | | |
| Db | 822 | KESKPSMIHOSPMPPTSOMMTAMESLKLSESGQTG--GPTVATGGPPQRAT'S-----QQM | 874 | | | |
| QY | 390 | VEQTTTTTTTTTTTTTARTVENKPNANTPAQGNVDTPESEDTMESRSSMASSTSTFFDTS | 449 | | | |
| Db | 875 | SRGATINSANMGASSGAAASATNQLSGAPSSSTGASSQOYHPKAPSSSSSST----- | 929 | | | |
| QY | 450 | SIGGPCRIRMLMKHRCMIIRRELLILRLFRWGIQISVYVSTIOHPPR---DVTNCA | 506 | | | |
| Db | 930 | --NPP-----HQQLTHNA-----SFSVTPSSYQIPTSTAVNVITSCM | 965 | | | |
| QY | 507 | RLGPNPSAGTQSTVARLALSGGLRHDWGG | 535 | | | |
| Db | 966 | PTSSSSSAFPNTPNRQTFHGKTEKDKGG | 994 | | | |
| RESULT | 8 | | | | | |
| G64887 | | probable tail fiber protein GP37 - Escherichia coli | | | | |
| C:Species: | Escherichia coli | | | | | |
| C:Date: | 12-Sep-1997 | #sequence_revision 17-Sep-1997 | #text_change 21-Jul-2000 | | | |
| C:Accession: | G64887; T09189 | | | | | |
| R:Blattner, F.R.; | Plunkett III, G.; | Bloch, C.A.; | Perna, N.T.; | Burland, V.; | Riley, M.; | |
| A.; | Rose, D.J.; | Mau, B.; | Shao, Y. | | | |
| Science | 277, 1453-1462, 1997 | | | | | |
| A:Title: | The complete genome sequence of Escherichia coli K-12. | | | | | |
| A:Reference number: | A64720; MUID:97426617 | | | | | |
| A:Accession: | G64887 | | | | | |
| A:Status: | nucleic acid sequence not shown; translation not shown | | | | | |
| A:Molecule type: | DNA | | | | | |
| A:Residues: | 1-1122 <BLAT> | | | | | |
| A:Cross-references: | GB:AE000234; GB:U000096; NID:g1787633; PIDN:AAC74454.1; PID:g17876 | | | | | |
| A:Experimental source: | strain K-12, substrain MG1655 | | | | | |
| R:Alba, H.; | Baba, T.; | Fujita, K.; | Hayashi, K.; | Inada, T.; | Isono, K.; | |
| A.; | Tomotura, K.; | Nakade, S.; | Nakamura, Y.; | Nashimoto, H.; | Nishio, Y.; | |
| moto, Y.; | Horiuchi, T. | | | | | |
| A:Title: | A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to t | | | | | |
| A:Reference number: | 216603; MUID:97251357 | | | | | |
| A:Accession: | T09189 | | | | | |
| A:Status: | preliminary; translated from GB/EMBL/DBJ | | | | | |
| A:Molecule type: | DNA | | | | | |
| A:Residues: | 3-1122 <AIB> | | | | | |
| A:Cross-references: | GB:AE000234; GB:U000096; NID:g1787633; PIDN:AAC74454.1; PID:g17876 | | | | | |
| Query Match | 5.1%; | Score 146.5; | DB 2; | Length 1122; | | |
| Best Local Similarity | 23.1%; | Pred. No. 0.55; | | | | |
| Matches | 92; | Conservative | 52; | Mismatches 163; | Indels 91; | Gaps 17; |
| QY | 79 | SEITLNDGFEVLH-----DHGPLDTLNROIGSSVFRVETOEDGKHIVGQ----- | 123 | | | |
| Db | 61 | SVILLVEGPPPPSHAGTITVYEDSQP-GTINDPLGAM-----TEDDARPEALRRLELMVEE | 114 | | | |
| QY | 124 | --RNGVETSVLSDQEYARLQSIDPEGKDKFVFTGGRGAGHAMVTVASDITEARQILRE | 181 | | | |

A32634
lactocipin (EC 3.4.21.96) precursor [similarity] - Lactococcus lactis subsp. cremoris
N:Alternate names: cell envelope-associated serine proteinase ptp.
C:Species: Lactococcus lactis subsp. cremoris
C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 21-Jul-2000
C:Accession: A32634
R:Vos, P.; Simons, G.; Sziegen, R.J.; de Vos, W.M.
J. Biol. Chem. 264, 13579-13585, 1989
A:Title: Primary structure and organization of the gene for a procaryotic, cell envelope-associated serine proteinase from *Lactococcus lactis* subsp. *cremoris*.
A:Reference number: A32634; MUID:89340435
A:Accession: A32634
A:Molecule type: DNA
A:Residues: 1-1962 <VOS>

Thu Sep 27 15:30:01 2001

Db 180 VTEASSAETTTAGTEASGEETTTSAVTEGSGEETTVAVVES---SGEPASSSTTSIP- 235
Qy 72 NPMRLAASEITLNDGFVLHDHGLPLTLNRQIGSSVFRVETQEDGKHIAVGQRNGVETSV 131
Db 236 -----TELSKND--QVTEASGE-ETI-----TAAATEASEETTTSAVTEGSGEDTTV 279
Qy 132 VL-----SDQEYARLQSIDPE-GKDKRVFTGGRGGAGHAMVTVASDITEARQRILELLEP 185
Db 280 VAVVELSGEQPASPSTSIPTELSKDDQV-----TEASGEETTTAAA--TEASE-----ET 327
Qy 186 KGTGESKGAGESKGVGELRESNCAENTTTQTSTSTS---SLRSDPKLWLALGTVATGL 242
Db 328 TTSVAVTEGSGEETTVAVVLESSG-----EPPASSSTSIPTELSKDDQVTEASGEETT-- 379
Qy 243 IGLAATGIVQALALTPDPDSPTTTDPPAAASATETATRDQLTKEAFQNPDPNQKNIDELG 302
Db 380 -----TAAATEASEETTT-----SAVTEGSGEDTTVVAVVLESSGEQPAS---SS 420
Qy 303 NAIPSGVLKDDV--ANIEQAKAAGEAKQAENNAQAQKYDEQQAQRQELKYSSG 360
Db 421 TSIPTELKDDQVTEASGEETTTAAATEASEETTS-----AVTEG 461
Qy 361 AGYGLSGALIL-----GGGIGVAVTAALHRKNQPV-----EQTTTTTTT-----TTTTSAR 406
Db 462 SGEDTTVVAVVLESSGEQPASPSTSIPTELSKDDQVTEASGEETTTAAATEASEETTTSAV 521
Qy 407 T-----VENKPANTPAGNVDPGSEDWESRRESSMASTSTFFDTS 450
Db 522 TEGSGEETTVAVVLESSGEEPASSSTSIPTELSKDDKVTEASGEETTTAAATDASS 577

Search completed: September 27, 2001, 14:21:47
Job time: 83 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:22:50 ; Search time 14.89 Seconds
(without alignments)
1286.019 Million cell updates/sec

Title: US-09-189-415a-4
Perfect score: 2851
Sequence: 1 MPIGNLGHNPVNNNSIPPAP.....SNSAVNTSNPPAGSHRPV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 146.5 | 5.1 | 1120 | 1 STFR_ECOLI | P76072 escherichia |
| 2 | 142 | 5.0 | 563 | 1 MUCS_BOVIN | P98091 bos taurus |
| 3 | 140.5 | 4.9 | 1026 | 1 STAU_DROME | P25159 drosophila |
| 4 | 136.5 | 4.8 | 461 | 1 US45_LACLC | P22865 lactococcus |
| 5 | 133.5 | 4.7 | 796 | 1 YS8A_CAEEL | Q09625 caenorhabdi |
| 6 | 132.5 | 4.7 | 1258 | 1 ICEN_ERWHE | P16239 erwinia her |
| 7 | 130.5 | 4.6 | 1034 | 1 ICEN_PANAN | Q47879 pantoea ana |
| 8 | 130 | 4.6 | 1077 | 1 HLES_DROME | Q02308 drosophila |
| 9 | 129.5 | 4.5 | 2411 | 1 DAB_DROME | P98081 drosophila |
| 10 | 129 | 4.5 | 784 | 1 SP4_HUMAN | Q02446 homo sapien |
| 11 | 129 | 4.5 | 2090 | 1 HFC1_MESAU | P51611 mesocricetu |
| 12 | 128.5 | 4.5 | 1902 | 1 P3P_LACLC | P15292 lactococcus |
| 13 | 127.5 | 4.5 | 3726 | 1 TRX_DROME | P20659 drosophila |
| 14 | 127 | 4.5 | 1337 | 1 DEXT_STRDO | P39653 streptococc |
| 15 | 126 | 4.4 | 997 | 1 BIR1_SCHPO | O14064 schizosacch |
| 16 | 125.5 | 4.4 | 1322 | 1 ICEA_PANAN | P20469 pantoea ana |
| 17 | 125 | 4.4 | 1694 | 1 IGAO_HAEIN | P44969 haemophilus |
| 18 | 125 | 4.4 | 1702 | 1 IGA2_HAEIN | P45384 haemophilus |
| 19 | 125 | 4.4 | 1772 | 1 MSP1_PLAYO | P13828 plasmodium |
| 20 | 125 | 4.4 | 1802 | 1 HKR1_YEAST | P41809 saccharomyc |
| 21 | 125 | 4.4 | 1902 | 1 P1P_LACLC | P16271 lactococcus |
| 22 | 124 | 4.3 | 1902 | 1 P2P_LACLC | P15293 lactococcus |
| 23 | 123.5 | 4.3 | 2483 | 1 PCX_DROME | P18490 drosophila |
| 24 | 123 | 4.3 | 1849 | 1 IGA4_HAEIN | P45386 haemophilus |
| 25 | 122.5 | 4.3 | 2842 | 1 APC_RAT | P70478 rattus norv |
| 26 | 122 | 4.3 | 2003 | 1 YDBA_ECOLI | P33666 escherichia |
| 27 | 122 | 4.3 | 2035 | 1 HFC1_HUMAN | P51610 homo sapien |
| 28 | 121.5 | 4.3 | 1079 | 1 IF2P_SCHPO | Q10251 schizosacch |
| 29 | 121.5 | 4.3 | 1189 | 1 YJH6_YEAST | P47035 saccharomyc |
| 30 | 121 | 4.2 | 1150 | 1 APMU_PIG | P12021 sus scrofa |
| 31 | 121 | 4.2 | 1306 | 1 MSB2_YEAST | P32334 saccharomyc |
| 32 | 120.5 | 4.2 | 746 | 1 7UP2_DROME | P16376 drosophila |
| 33 | 120 | 4.2 | 797 | 1 VGLX_HSVB | P28968 equine herp |

RESULT 1

| ID | STFR_ECOLI | STANDARD; | PRT; | 1120 AA. |
|----|--|-----------|------|----------|
| AC | P76072: P77560; | | | |
| DT | 01-OCT-2000 (Rel. 40, Created) | | | |
| DT | 01-OCT-2000 (Rel. 40, Last sequence update) | | | |
| DT | 01-OCT-2000 (Rel. 40, Last annotation update) | | | |
| DE | SIDE TAIL FIBER PROTEIN HOMOLOG FROM LAMBDOID PROPHAGE RAC. | | | |
| GN | STFR OR B1372. | | | |
| OS | Escherichia coli. | | | |
| OC | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; | | | |
| OC | Escherichia. | | | |
| OX | NCBI_TaxID=562; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-K12 / MG1655; | | | |
| RA | MEDLINE=97426617; PubMed=9278503; | | | |
| RA | Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., | | | |
| RA | Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., | | | |
| RA | Gregor J., Davis N.W., Kirkpatrick H., Goeden M.A., Rose D.J., | | | |
| RA | Mau B., Shao Y.; | | | |
| RT | "The complete genome sequence of Escherichia coli K-12."; | | | |
| RL | Science 277:1453-1474(1997). | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN-K12; | | | |
| RA | MEDLINE=97251357; PubMed=9097039; | | | |
| RA | Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T., | | | |
| RA | Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M., | | | |
| RA | Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K., | | | |
| RA | Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N., | | | |
| RA | Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J., | | | |
| RA | Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.; | | | |
| RT | "A 570-kb DNA sequence of the Escherichia coli K-12 genome | | | |
| RT | corresponding to the 28.0-40.1 min region on the linkage map."; | | | |
| RL | Corresponding to the 28.0-40.1 min region on the linkage map."; | | | |
| CC | -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announcement/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; AE000234; AAC74454.1; ALT_INIT. | | | |
| DR | EMBL; D90774; BAA14966.1; -. | | | |
| DR | EMBL; D90775; BAA14975.1; -. | | | |
| DR | EcoGene; EG13370; STFR. | | | |
| DR | HSSP; P04002; 1WFA. | | | |
| DR | InterPro; IPR000122; - | | | |
| DR | Hypothetical protein; Fiber protein; Repeat. | | | |
| SW | SEQUENCE 1120 AA; 113779 MW; 542E59D71EE795B4 CRC64; | | | |

ALIGNMENTS

| | | | | | | |
|----|-------|-----|------|---|------------|--------------------|
| 34 | 119 | 4.2 | 484 | 1 | P60_LISMO | P21171 listeria mo |
| 35 | 119 | 4.2 | 1048 | 1 | P100_RCMTA | P08318 human cytom |
| 36 | 119 | 4.2 | 1581 | 1 | PPRB_HUMAN | Q15648 h peroxisom |
| 37 | 119 | 4.2 | 1603 | 1 | PSC_DROME | P35820 drosophila |
| 38 | 118.5 | 4.2 | 516 | 1 | P54_ENTFC | P13692 enterococcu |
| 39 | 118 | 4.1 | 662 | 1 | MUC1_XENLA | Q05049 xenopus lae |
| 40 | 118 | 4.1 | 1781 | 1 | AKAC_HUMAN | Q02952 homo sapien |
| 41 | 117.5 | 4.1 | 523 | 1 | P60_LISSE | Q01838 listeria se |
| 42 | 117.5 | 4.1 | 1025 | 1 | SLAP_CAUCR | P35828 caulobacter |
| 43 | 117 | 4.1 | 567 | 1 | CHI3_CANAL | P40954 candida alb |
| 44 | 117 | 4.1 | 681 | 1 | VGP_MABVP | P35254 marburg vir |
| 45 | 117 | 4.1 | 1036 | 1 | NIT2_NEUCR | P19212 neurospora |

Query Match 5.1%; Score 146.5; DB 1; Length 1120;
Best Local Similarity 23.1%; Pred. No. 0.39;
Matches 92; Conservative 52; Mismatches 163; Indels 91; Gaps 17;

QY 79 SEITLNDGFEVLH-----DHGPIDLTLNRQIGSSVFRVETQEDKHIAYGQ-----123
DB 59 SVILLVEGPPSHAGTITVYEDSQP-GTLNDFLGAM-----TEDDARPEALRRFELMVEE 112
QY 124 --RNGVETSVLSDOEVARLOSIDPEKDKFVTGGRGGAGHAMVTVASDITEARQRILE 181
DB 113 VARN---ASAVAQNTAAKKSADAS-----TSAREATHA--ADAUSARA-----154
QY 182 LLEPKGTGSGKAGSGVGELESNSGAENTTETQSTSTSLRSDPKLWLALGVATG 241
DB 155 -----ASTSAGQAASSAQSSAGTATKATASKSAAAEESK-----SAAAT 199
QY 242 LIGLAATGIVOALA----LPEPDSPTTDPDAAASATETATRDQLTKAFQNPDKQVN 297
DB 200 SAGAAKTSETNASLSQAASATATTKASEAATSARDAAA-----SKEAKSSET----251
QY 298 IDELGNAIPSGVLKDDVVANIEEQAKAAGEAK-QQAIENNAQAQKYDEQQAQKQBEEL 356
DB 252 -----NASS-----ASSAASATAGNSAKAAKTSETNARSSETAAGSASAAAGSK 299
QY 357 VSGAGYGLSGALILGGIGVAVTAALHRKNQPVQETTTTTTTTTTTTTTTSARTVENKPNANT 416
DB 300 --TAAASASAASTAGSAGASATAA---GKSAESAASASTATTKAGEATEQAASAAAR 353
QY 417 PAQ-----GNVDTPGSEDTMESRRSSMASTSTFTDTS 450
DB 354 SASAAKTSETNAKASETSAESSKTAASASSASSASSAS 391

RESULT 2
MUCS_BOVIN STANDARD; PRT; 563 AA.
AC P98091;
DR 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE SUBMAXILLARY MUCIN-LIKE PROTEIN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Submaxillary gland;
RX MEDLINE=90370871; PubMed=2204065;
RA Bhargava A.K., Woitach J.T., Davidson E.A., Bhavanandan V.P.;
RT "Cloning and cDNA sequence of a bovine submaxillary gland mucin-like
protein containing two distinct domains";
RL Proc. Natl. Acad. Sci. U.S.A. 87:6798-6802(1990).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC -1- SIMILARITY: TO PORCINE APOMUCIN.
CC -1- SIMILARITY: CONTAINS 1 VNFC DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M36192; AAA30657.1; -
DR HSSP; P00214; 1FTC.
DR InterPro; IPR000359; -
DR InterPro; IPR001007; -
DR Pfam; PF00007; Cys_knot; 1.

DR PROSITE; PS01208; VNFC; 1.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
KW Repeat; Glycoprotein.
FT DOMAIN 61 158
FT REPEAT 61 71
FT REPEAT 112 122
FT REPEAT 148 158
FT REPEAT 338 404
FT DOMAIN 471 555
FT SIMILAR 1 282
FT DISULFID 471 518
FT DISULFID 485 532
FT DISULFID 494 548
FT DISULFID 498 550
FT DISULFID 7 554
FT CARBOHYD 28 28
FT CARBOHYD 394 394
FT CARBOHYD 477 477
SQ SEQUENCE 563 AA; 58913 MW; AB326CD78E5FFFCF CRC64;

Query Match 5.0%; Score 142; DB 1; Length 563;
Best Local Similarity 20.3%; Pred. No. 0.29;
Matches 75; Conservative 48; Mismatches 169; Indels 78; Gaps 13;

QY 88 EVLHDHGPLDLNRQIGSSVFRVET-----QEDGKHIAGVGRNGVETSVLS 134
DB 2 KVLQENSRAHS---GSSHTAATLLIVSNSTSGTLRPEDNTAVAGGQATGRVT-----53
QY 135 DOEVARLOSIDPEKDKFVTGGRGGAGHAMVTVASDITEARORILELLEPKGTGSKGA 194
DB 54 -----GTTK-VIPGTTVAPGSSNTSTSLGESRTRIGRI-----TGATTGT 94
QY 195 GESKGVGELRESNSGA-ENTTETQSTSTSLRSDPKLWLALGVATFGLGLAATGIVQA 253
DB 95 SKRSSPGS-KTNGTALSGTTPVAPSSNTGATTS-----LGSGETTGGIKIVTMGVTTG 148
QY 254 LALTPEPDSPTTDPDAAASATETATRDQLTKAFQNPDKQVNIDELGNAPSGVLKDD 313
DB 149 TTIAPGSSNTKATTPTEVRTTETATTT-----TSRHSSDATGSGIOFTGI-----196
QY 314 VVANIEQAQAKAAGEAKQAQAIENNAQAQKYDEQQAQKQBEELKVSSGAGYGLSGALILGG 373
DB 197 -----TGCGSGTTPSPGFGNAPATTFKEH-----VRTTETRLSGTTRGSGTTYPE 244
QY 374 GIGVAVTAALHRKNQPVQETTTTTTTTTTTTTTTSARTVENKPNANTPAQNVDPGSEDTMES 433
DB 245 SSNTGTSTGVGR-----QTSTAVVSGRVTGV-SESSSPGTSKEASETTTGPGISSTGTST 297
QY 434 RRSMASTSS 443
DB 298 SKSNRITSS 307

RESULT 3
STAU_DROME STANDARD; PRT; 1026 AA.
ID STAU_DROME
AC P25159;
DR 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE MATERNAL EFFECT PROTEIN STAUFEN.
GN STAU.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91300552; PubMed=1712672;
RA St Johnston D., Beuchle D., Nusslein-Volhard C.;

Query Match 4.8%; Score 136.5; DB 1; Length 461;
Best Local Similarity 19.1%; Pred. No. 0.47; Indels 123; Gaps 14;
Matches 84; Conservative 69; Mismatches 163; Indels 163; Gaps 14;

QY 181 ELLEPKGTGESKAGSGKVGELRESNGAENTTQTSTSSLSRSDPKLMLALGTAT 240
DB 32 DIAKQATISSAQAQAQAVDSLSQKVSLSQKQOTSTKAQIAKIESELKALNAQIAT 91
QY 241 GLIGL-AATGIVQALALTPEDSPPTTDDPA-----AASATETATRDOL-- 283
DB 92 LNESIKERTKLEAQAARSQAQVNSATNMDAVVNSKSLTDVIOKVTAIATVSSANKQILE 151
QY 284 -----KFAQNPNDKNVNDI--LGNALPS-----GVLKDDVVANIEEQAKAAG 326
DB 152 QQKEQKELSKQSETVKKVNFVLSQSLDSQAQELTSSQAELK---VATLNYQATIA 208
QY 327 EEAQQAIAENNAQAQKYDQQAQKQOE--ELKYSSGAGYGLSGLILGGIGVAVTAALH 384
DB 209 AODKQALLDEKAAEAQAQAQAAKQAAEAQAQAQAQAQAQAQAQAQAQAQAQAQA 268
QY 385 RKQNPVQQTSTTTTTSARIVENKPAANTPAQGNVDTPGSEDTHMESRRSSMASTST 444
DB 269 SSSQAPQVSTSTQNTSNASNSNSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 324
QY 445 FFDSSIGGPCIRMLKHKRCMIRRCRLILIRLFRINGIQISVVYSTIQHPPTDTDN 504
DB 325 GTSTGNTGG-----TGTG 337
QY 505 GARL-----LGNPSAG-----IQSYARLALSG-----GLRHDG-- 534
DB 338 GSGINSSPIGNPYAGGCTDYVMOYFAAQGIYIRNIPGNGGQWASGPAQGVHLVYGAA 397
QY 535 -GLTGGNSA--VNTSNP 550
DB 398 PGVIASSFADFVGYNAP 416

RESULT 5
YS8A_CAEEL STANDARD; PRT; 796 AA.
ID YS8A_CAEEL
AC Q09625;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 84.3 KDA PROTEIN ZK945.10 IN CHROMOSOME II.
GN ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wilkinson-Sproat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z48582; CAA88469.1;
DR EMBL; Z48544; CAA88469.1; JOINED.
DR EMBL; Z48544; CAA8844.1;
DR EMBL; Z48582; CAA8844.1; JOINED.
DR WormPep; ZK945.10; C01732.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 11 30 POTENTIAL.
FT DOMAIN 273 546 SER/THR-RICH.
FT DOMAIN 636 732 SER/THR-RICH.

SQ SEQUENCE 796 AA; 84306 MW; 76DC5B03E6357A6A CRC64;
Query Match 4.7%; Score 133.5; DB 1; Length 796;
Best Local Similarity 19.7%; Pred. No. 1.4;
Matches 82; Conservative 76; Mismatches 176; Indels 83; Gaps 14;

QY 68 GLPVNPMRLAASEITLND-----GFEVLHDHPLDLNRLQIGSSVFRVETQEDGKHI 119
DB 137 GLFLNSTWITLNEVNDDEISIAVEAKYEVGYDDG----IDRCDSGLMW---LQVGGNEM 189
QY 120 A-VGQNGVETSVLSQYARLQSIDPEKQKFEVETGGRGGAGHAMVTVASDITEARQR 178
DB 190 ALIGYREKCESEI--NEEYARMKRPRSEK-----STAISDSQGV 230
QY 179 ILELLEPKGTGESKAGSGKVGELRESNGAENTTQTSTSSLSRSDPKLMLALGTAT 238
DB 231 YDGVQVLKGVRAKQFQSMRTSGSPTLRMRKRDAGDNTCDYTIESTSTSTTTTPTTSTV 290
QY 239 -ATGLIGLAATGIVQALALTPEDSPPTT-----DPDAAASATETATRDQLTKEAFQ 289
DB 291 TSTTTVPTSTSTVTAMST 350
QY 290 NPDNQVNDELGNALPSGVKLD--DVANIERQAKAAG--EEAKQQAIAENNAQAQKYDE 346
DB 351 SPSS-----TTLSTSTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 405
QY 347 QQAKRQBELKVGAGYGLSGALILGGIGVAVTAALHRRKNQVPEQTTTTTTTTTTT--S 404
DB 406 EPSTSTTTTEVTS-----TSSVTVTTEPTTPTTPTTPTTPTTPTTPTTPTTPTT 445
QY 405 ARTVENKPA-----NNTPAQGNVDTPGSEDTHMESRRSSMASTSTSTSTSTSSIG 452
DB 446 TSTVTTSPSTVSTVSTVSTVSTVSTVSTVSTVSTVSTVSTVSTVSTVSTVSTVST 502

RESULT 6
ICEN_ERWHE STANDARD; PRT; 1258 AA.
ID ICEN_ERWHE
AC P16239;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ICE NUCLEATION PROTEIN.
GN ICEE.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwinia.
OX NCBI_TaxID=557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M1;
RX MEDLINE=90152370; PubMed=2515997;
RA Warren G.J., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
RT herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
RT Gene 85:239-242(1989).
RL CC
CC -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T. FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; M26382; AAA24823.1; -.
DR PIR; J00188; J00188.
DR HSRP; P06620; LINA.
DR InterPro; IPR000258; -.
DR Pfam; PF00818; Ice_nucleation; 65.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 45.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1258 AA; 125084 MW; 590E8A130077EBD4 CRC64;

Query Match 4.6%; Score 132.5; DB 1; Length 1258;
Best Local Similarity 20.4%; Pred. No. 2.8; Mismatches 226; Indels 141; Gaps 20;
Matches 110; Conservative 62;
QY 26 TDGAGGRGLI---NSTGLGSRALFTPVNRSMADSGDNRAVDVPLPVPNMLAASEI- 81
DB 133 TETAGDSSTLAGYSGTGTAGADSTLVAGYSGTQAGE-ESSQAGYSGTQTMKGSDLT 241
QY 82 -----TLNDGFEVLHDHGPLDT-----LNRTGSSVFRVETQEDGKHIAVGORN--- 125
DB 242 AGYSGTGTAGDSSSLIAGYSGTQAGEDSSLTAGYSGT---QTAQKGSDLTLAGYSGTGT 297
QY 136 -GVETSVVLSDOEYARLQSIDPE-----GKDKFVTGGRGGAG----- 162
DB 298 AGADSSLIAG---YGSTQAGEESTQTAGYSGTQTAQKGSDLTLAGYSGTGTAGDSSSLIA 354
QY 163 -----HAMVTVASDITEARORILELLEPKGTGSKGAGESKVGCELRESNSGAEN 212
DB 355 GYGSQTQAGEDSSSLTLAGYSGTQTAQKGSDLTLAGYSGTGTAGADSSLIAGYSGTQAGEES 414
QY 213 T-TETQSTSTSSLRSDPKMLALGTAVATGLIG-----LAATGIVQALALTPEDSPPT- 265
DB 415 TOTAGYSGTQTAQKGS-----LTAGYSGTGTAGDSSSLIAGYSGTQ---TAGEDSSSLTA 466
QY 266 -----TDPDAASATETATRDQLTKEAF-----QNPDKNQK 295
DB 467 GYGSQTQAKGSDLTLAGYSGTSTAGYESLIAGYSGTGTAGYSGTTLTLAGYSGTQTAQNES 526
QY 296 VNIDELGNAIPSGVLKDDVAVANIEQAQA-----AGEAKQAQAIENNAQAQKRYDEQ 355
DB 527 DLITCYGSTSTAGA-----NSSLIAGYSGTQTASVNSLTAGYSGTQTAQREGSDLT 576
QY 356 KVSAGYGLSGAL-----ILGGIGVAVTAALH-----RKNQPVETQTTTTTT 399
DB 577 T-----AGYSGTGTAGSDSSIIAGYSGTQTASVHSSLTLAGYSGTQTAQREGSVLTTCYGSTS 632
QY 400 TTTTSARTVENKPNANTPAQGNVDPGSEDIMESRRS-----MASTSTFTDSSIGG 453
DB 633 TAGADSSLIAGYSGTQTAGYNSILTAGYSGTQTAQEGSDLTLAGYSGTSTAGADSSLIAG 691

RESULT 7

ICEN_PANAN STANDARD; PRT; 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ICE NUCLEATION PROTEIN INAU.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KUIN-3;
RX MEDLINE=94264407; PubMed=7764866;

RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
RT uredovora";
RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; D14992; BAA03636.1; -.
DR HSRP; P06620; LINA.
DR InterPro; IPR000258; -.
DR Pfam; PF00818; Ice_nucleation; 51.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 34.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1034 AA; 103378 MW; FA22523D333EADD CRC64;

Query Match 4.6%; Score 130.5; DB 1; Length 1034;

Best Local Similarity 20.9%; Pred. No. 2.9; Mismatches 229; Indels 139; Gaps 23;
Matches 114; Conservative 64;

QY 26 TDGAGGRGLI---NSTGLGSRALFTPVNRSMADSGDNRAVDVPLPVPNMLAASEI- 81
DB 133 TETAGDSSTLAGYSGTGTAGDSTLVAGYSGTQAGE-ESSQAGYSGTQTMKGSDLT 241
QY 82 -----TLNDGFEVLHDHGPLDT-----LNRTGSSVFRVETQEDGKHIAVGORN--- 125
DB 242 AGYSGTGTAGDSSSLIAGYSGTQAGEDSSLTAGYSGT---QTAQKGSDLTLAGYSGTGT 297
QY 136 -GVETSVVLSDOEYARLQSIDPE-----GKDKFVTGGRGGAG----- 162
DB 298 AGADSSLIAG---YGSTQAGEESTQTAGYSGTQTAQKGSDLTLAGYSGTGTAGDSSSLIA 354
QY 163 -----HAMVTVASDITEARORILELLEPKGTGSKGAGESKVGCELRESNSGAEN 212
DB 355 GYGSQTQAGEDSSSLTLAGYSGTQTAQKGSDLTLAGYSGTGTAGADSSLIAGYSGTQAGEES 414
QY 213 T-TETQSTSTSSLRSDPKMLALGTAVATGLIG-----LAATGIVQALALTPEDSPPT- 265
DB 415 TOTAGYSGTQTAQKGS-----LTAGYSGTGTAGDSSSLIAGYSGTQ---TAGEDSSSLTA 466
QY 266 -----TDPDAASATETATRDQLTKEAF-----QNPDKNQK 295
DB 467 GYGSQTQAKGSDLTLAGYSGTSTAGYESLIAGYSGTGTAGYSGTTLTLAGYSGTQTAQNES 526
QY 296 VNIDELGNAIPSGVLKDDVAVANIEQAQA-----AGEAKQAQAIENNAQAQKRYDEQ 348
DB 527 DLITCYGSTSTAGA-----NSSLIAGYSGTQTASVNSLTAGYSGTQTAQREGSDLTLAGYSGTGT 585
QY 349 AKRQELKVSQ-----AGYGLSGALILGGIGVAVTAALH-----RKNQPVETQTTTTT 397
DB 586 AQENSDLTTCYGSTSTAGYDSS-----LIAGYSGTQTAGVHSLTLAGYSGTQTAQERSDLT 641
QY 398 T---TTTTTART---VENKPNANTPAQGNVDPGSEDIMESRRS-----MASTSTFTD 447
DB 642 TCYGSTSTAGADSSLIAGYSGTQTAGYNSILTAGYSGTQTAQENSDLTTCYGSTSTAGYE 701

Best Local Similarity 18.9%; Pred. No. 3.2;
Matches 145; Conservative 66; Mismatches 244; Indels 312; Gaps 29;

QY 448 TSSIG 453
Db 702 SSIAG 707

RESULT 8
HLES_DROME STANDARD; PRT; 1077 AA.
AC Q02308;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE HAIRLESS PROTEIN.
GN H.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92387549; PubMed=1516831;
RA Bang A.G., Posakony J.W.;
RT "The Drosophila gene Hairless encodes a novel basic protein that
controls alternative cell fates in adult sensory organ development.";
RL Genes Dev. 6:1752-1769(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93041287; PubMed=1419850;
RA Maier D., Stumm G., Kuhn K., Preiss A.;
RT "Hairless, a Drosophila gene involved in neural development, encodes
a novel, serine rich protein.";
RL Mech. Dev. 38:143-156(1992).
CC -!- FUNCTION: IS A POTENT ANTAGONIST OF NEUROGENIC GENE ACTIVITY
DURING SENSORY ORGAN DEVELOPMENT. THE EXPRESSION OF DISTINCT CELL
FATES BY THE TRICHOGEN (SHAFT) / TORMOGEN (SOCKET) SISTER CELL
PAIR DEPENDS ON THE LEVEL OF H ACTIVITY. A CERTAIN THRESHOLD LEVEL
OF H ACTIVITY IS REQUIRED, BELOW WHICH BOTH SISTER CELLS ADOPT THE
TORMOGEN FATE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- TISSUE SPECIFICITY: OVARY, EMBRYOS, LARVAL AND PUPAL IMAGINAL
DISCS.
CC -!- SIMILARITY: CONTAINS A "PRD MOTIF".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; M95192; AAA28607.1; ALT_INIT.
DR EMBL; X67239; CAA47664.1; -.
DR HSSP; P04002; 1WFA.
KW FlyBase; FBgn0001169; H.
KW Developmental protein; Nuclear protein; DNA-binding.
FT DOMAIN 115 123 THR-RICH.
FT DOMAIN 642 648 POLY-SER.
FT DOMAIN 879 891 POLY-ALA.
FT DOMAIN 937 946 POLY-ALA.
FT DOMAIN 964 974 ALA-RICH.
FT DOMAIN 979 1008 HIS/PRO-RICH (PRD MOTIF).
FT CONFLICT 151 151 S -> A (IN REF. 2).
FT CONFLICT 702 703 OH -> LL (IN REF. 2).
FT CONFLICT 891 891 A -> R (IN REF. 2).
FT CONFLICT 964 967 RAVA -> RLLP (IN REF. 2).
FT CONFLICT 974 974 MISSING (IN REF. 2).
SQ SEQUENCE 1077 AA; 111039 MW; A94BFIA27579E2F1 CRC64;

Query Match 4.6%; Score 130; DB 1; Length 1077;

RESULT 9

DAB_DROME STANDARD; PRT; 2411 AA.
ID DAB_DROME
AC P98081;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DISABLED PROTEIN.
GN DAB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93194063; PubMed=7680635;


```

DR PRINTS: PR000048; ZINC_FINGER.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT DOMAIN 7 11
FT DOMAIN 12 19
FT DOMAIN 122 130
FT DOMAIN 185 188
FT DOMAIN 647 729
FT ZN_FING 647 729
FT ZN_FING 677 701
FT ZN_FING 707 729
FT CONFLICT 197 197
FT CONFLICT 379 380
FT CONFLICT 386
FT SEQUENCE 784 AA; 82025 MW; 3C4EAE28CB81FB CRC64;

Query Match 4.5%; Score 129; DB 1; Length 784;
Best Local Similarity 20.0%; Pred. No. 2.5;
Matches 112; Conservative 71; Mismatches 234; Indels 144; Gaps 21;

QY 3 IGNLGHNPVNNNS---IPAPPL-----PSQ-----TDAGRGQLINSTGPIG-SRAL 47
DB 58 IGTPEGNATGQQIITDPSQGLVQLNQPOQLVLTQLAGNAWOLVASTPPASKENNV 117
QY 48 FTPVRNMAISGDNRAISDVGLPNPMLRAASEITLNDGFVLHDHGPDLTLNRQIGSV 107
DB 118 SQAPASSSSSSSSNNNGS-----ASPTKSGNSSTPGQFQVQLVQNPSSGVQVPIQL 171
QY 108 FRVETQEDGKHIAVGQNGVETSVVLSDQYARLQSDPEGDKFVFT-GRGGAGHAMV 166
DB 172 QTVEGQI-----QINPTSSSLQDLQGIKLS---AGNNQAULTAARTASGNILA 221
QY 167 -----TVASDITEARQILLELLEPKGT-----GESK 192
DB 222 QNLANTVPVQIRPGVSIPLQLTLPQTQAQVVTLPINIGGVTLALPVINNAAGGGTG 281
QY 193 GAGESKGVGLRESNGAENTETQTSLSRLSDPKLWALGTAT-----GLIGL 245
DB 282 QVGPAATADSGTNGNLVSTPTNTTASTMPESSTTCTTASTSLTSSDTLVSS 341
QY 246 AATGIVQALALTPEDSTPTTDPDAASATETATR-DQLTKEAFONPDNOKVNIDELG- 302
DB 342 ADTG---QYASTSASSERTIEESQTPAATESAQSSQLHANGMONGQDQNSLQVQI 398
QY 303 -----NAIP-----SG-----VLKDDVWANIERQA----- 322
DB 399 VQPILOIQIQPOQOQIIQAIIPQSFQSQGTITQIQOQLQNVQLQVNPQVLIRA 458
QY 323 ---KAAGEAKQQALENNAQAKKYDEQAKRQELKYSAGYGLSALLIGGGIGVAV 379
DB 459 PLTSPGQISQMTQVQVQNIQSLNQVQAGLSQQLTITPVSS---SGGTLAQIAPVAV 515
QY 380 TAAHKKNPQVEQTITTTTTTTSARVENKPNANTPAQGNVDTPGS-----EDTME 432
DB 516 AGA-----PITLNTAQLASVNLQTSVANLGAAGVQVQGVPTTITSVAGQQQGGGVK 569
QY 433 SRSSMASTSSFTFTSSIGG 453
DB 570 VQQTATAPV-----TVAVGG 584

RESULT 11
HFC1_MESAU
ID HFC1_MESAU STANDARD; PRT; 2090 AA.
AC P51611;
DT 01-OCT-1996 (Rel. 34, Created)
DE 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE HOST CELL FACTOR C1 (HCF) (VP16 ACCESSORY PROTEIN) (HFC1) (VCAF)
DE (CFF).

GN Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_taxid=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Goto H., Nishitani H., Umene K.I., Nakabepu Y., Nishimoto T.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY
CC SIMILARITY).
CC SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
CC 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N-
CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
CC NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).
CC SUBCELLULAR LOCATION: NUCLEAR.
CC DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC SIGNAL (BY SIMILARITY).
CC PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC AT A DEFINED SITE, PPCE/THET, WITHIN THE HCF REPEAT (BY
CC SIMILARITY).
CC SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; DA5419; BAA08258.1;
CC HSSP; P02751; 1FNA.
CC InterPro; IPR001777;
CC Pfam; PF00041; fn3; 2.
CC Nuclear protein; Repeat.
CC REPEAT 44 89
CC FT REPEAT 93 140
CC FT REPEAT 148 194
CC FT REPEAT 217 265
CC FT REPEAT 266 313
CC FT DOMAIN 1010 1448
CC FT REPEAT 1010 1035
CC FT REPEAT 1072 1097
CC FT REPEAT 1101 1126
CC FT REPEAT 1157 1182
CC FT REPEAT 1295 1320
CC FT REPEAT 1323 1348
CC FT REPEAT 1358 1383
CC FT REPEAT 1423 1448
CC FT SEQUENCE 2090 AA; 214942 MW; E495E8B1F2385E17 CRC64;

Query Match 4.5%; Score 129; DB 1; Length 2090;
Best Local Similarity 19.3%; Pred. No. 8.6;
Matches 121; Conservative 71; Mismatches 224; Indels 224; Gaps 27;

QY 3 IGNLGHNPVNNNSIPAPPLPSQTDGAGRG---QLINSTGPLGSRALFTPVRNMAISG 59
DB 677 ISNLG---KVMVSVQTKPVQISAVTQASTGPVTQIIQKGPLPAGTIL-----KLVTSA 728
QY 60 DNRAISV-----PGLPVNPMPLAASEI---TLNDGFVLHDHGPDLTLNRQIGSVFV 110
DB 729 DGKPTTIITTTQASGAGSKPTILGISSVSPSTKPGTTTIKTIPMSAIIQAGAT---- 784
QY 111 ETQEDGKHIANQGVNVTSSVLSQVY-----ARLQSIDP-----EGDKFVVF 154

```


Db 785 -----GVTSTPGIKSPITIIITKVTMTSGTGAPAKIITAVPKIATCHGQGVTVVVL 835
QY 155 TGGRGAGHAMVTV-----ASDIT-EARORILELLEPKGTGESKGAGSGVGELES 206
Db 836 KGAPQGAIIIRVPMGSRVLPVTVSAVPAVTVLVVKGTTGTTGLTGTGTVSTSLA 895
QY 207 NSGABNTTETQSTSTSLRSDPKLWALGATVATGLIGLAATGIVQALALPEPDSPTTT 266
Db 896 GAGAHSTASLATPIT-----LGTIAT-----LSQVINPAITV----- 931
QY 267 DPDAASATEATRDQLKEAFQFNDKQVNDLGNAPISGVLKDDVVANIEQAQKAA 326
Db 932 ---SAAQTLLTAAGLTPTTITMOPVSOPTQVTLI--TAPSGV----- 969
QY 327 BEAKQAIENNAQAKKYD-----EQAKQEEELKVS-----SGAGYGLSGALLGGGICV 377
Db 970 -----EQPVHDLVSVILASPTTQPTATVTVIADSGGQGVPG----- 1007
QY 378 AVTAALHRKNOPVEQTTTTTTTTTTTTTARTVENKPAANTPAQ----- 419
Db 1008 ---TVTLVCSNPPCETHETGTTTATTT--VVANLGGHPQPTQVQVFCVDRQEAASLVTS 1063
QY 420 -----GNVDTGSEDTMESRSSMASTSTFTDSSIGGPCIRMLKLMKHRMIRRCRL 474
Db 1064 VQQNGNVVRCNSPPCETHETGTGTATT--ATSNMAG-----QHGC----- 1104
QY 475 ILIRLFRWIGIISVYTIQHPPTDNGARLLGNPSAGIQSTYARLALSGGLRHDWG 534
Db 1105 -----SNPCETHETGT-----TSTATTAMS-----SMG 1128
QY 535 GLTGGSNAVNTSNPP-----APGS 555
Db 1129 ---TGOQDRHTSSNPTVVRITVARGA 1153
RESULT 12
P3P_LACLC
ID P3P_LACLC STANDARD; PRT; 1902 AA.
AC P15292;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PIII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOGEPIN) (CELL WALL-
DE ASSOCIATED SERINE PROTEINASE).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Plasmid.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC STRAIN=SK11;
RX MEDLINE=89304035; PubMed=2760036;
RA Vos P., Simons G., Siezen R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
RT cell envelope-located serine proteinase."
RL J. Biol. Chem. 264:13579-13595(1989).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J04962; AAA03533.1; ALT_SEQ.

PIR: A32634; A32634.
DR HSP: P00782; 2SEF.
DR MEROPS: S08.019; .
DR InterPro: IPR000209; .
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF00082; peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
FT SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;
Query Match 4.5%; Score 128.5; DB 1; Length 1902;
Best Local Similarity 22.7%; Pred. No. 8.1;
Matches 105; Conservative 56; Mismatches 197; Indels 105; Gaps 18;
QY 36 INSTGPLGSRALFTVPR-----NSWADSGDNRAVDVPGLPVNPWR-LAASEITLND 85
Db 1452 INSGKP-GHMAIDQPVKLEGGKNTLVAVTDESDNTTNNITVYVYEPKTLAAPTPTS- 1509
QY 86 GFEVLHDHGPDLTLNRQIGSSVFRVETOEDGKHIAVGORNGVETSVVLSQDQYARLQSID 145
Db 1510 -----TTEPACTVTLTANAAA---TGFTVQYSADGGKTYQDVP-----AAGVTIT 1551
QY 146 PEGKDKFVTGGRGAGHAMVTVASDI-----TEARQRIELELLEPKGTGESKGAGES 197
Db 1552 ANGTEFKFKSTDLGYNESPAVDVYVTVNIRKADDPQAQQAQKQELTNLIASAKTILSASGKYDD 1611
QY 198 KGVGELRESNSGAENTTETOTSTSSLR-SDPKLWALGTVATGLIGLAATGIVQALAL 256
Db 1612 ATTTALAATOKAQ--TALDQTNASVDSLTGANRDLQTAINQLAAKLPADKKTSLNLQLOS 1670
QY 257 TPE---PDSPTTTPDAAAATETATRDQLTKEA---FQNPQNKVNIDELGNALPSGYL 310
Db 1671 VKDALGTDLGNQTDTP--STGKTTFTAAALDDLVAQAQAGTQTDQLQATLAK-----I 1719
QY 311 KDDVVANIEEQAKAGE---EAKQAI-----ENNAQAKKYDEQQAQR 351
Db 1720 LDEVLAKEAGIEKAATPAEVGNKADAATGKTWYADIADTLTSGQASADASDKLAHLQALQ 1779
QY 352 QEELKVSS-----GAGYGLSGALLGGGIGVAVTA-----ALHRKNOP 389
Db 1780 SUKTVAAAVEAAKTVGKGDGTTGSDGGGQGTAPAPADPGTGKDKGDEGSGPSSGGNIP 1839
QY 390 VEQTITTTTTTTTTTSARTVENKPAANTPAQGNVDTPGSEDTME 432
Db 1840 TKPATITTTTDDTTDR-----NGQLTSGKGLPKTGETTE 1875
RESULT 13
TRX_DROME
ID TRX_DROME STANDARD; PRT; 3726 AA.
AC P20659; Q27255; Q27327;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRITHORAX PROTEIN.

Thu Sep 27 15:30:03 2001

```
DR PROSITE: PS01282; BIR_REPEAT_1; FALSE_NEG.
DR PROSITE: PS50143; BIR_REPEAT_2; 2.
KW Cell Division; Repeat.
FT REPEAT 25 99 BIR 1.
FT REPEAT 120 194 BIR 2.
FT DOMAIN 80 83 POLY-ASP.
FT DOMAIN 312 319 POLY-ASP.
FT DOMAIN 487 490 POLY-SER.
SQ SEQUENCE 997 AA; 112579 MW; 952A6BAFA5C489FA CRC64;

Query Match 4.4%; Score 126; DB 1; Length 997;
Best Local Similarity 19.8%; Pred. No. 5;
Matches 98; Conservative 62; Mismatches 205; Indels 130; Gaps 17;

QY 13 NNSIPPAPPLPSQTDGAGRGQINGTGLGSRALFTPVNSMADSGDNRASDVPGLPVN 72
Db 222 DNSILPVSPTRDSTKSHKTLNFSRKNLNLARPLTMSLYTNTSEKDSQPTRAPSPK 281
QY 73 PMRLAASEITLNDGFEVLHDHGLDILNRQIGSS-----VFRVETQED----- 115
Db 282 PVLLTAPR-----RKNKSPKSKPAVKPKPIFSDDEDDDLTASQ 324
QY 116 -----GKHIAVGQRNGVETSVVLSQDEYARLQSIDPEGKDKFVFTGGRGAGHAMTV 168
Db 325 PFSKGICNDSMQVAKKNFTETIPLKDEKNELEHLVSPATS--VHTTVSDITGHOSVTD 382
QY 169 ASDITEARQRIELLEPKGTGSGKAGESKVGELRESNSGA-----ENTTETQTST 222
Db 383 ESD-----EQNNCMSTPPKITEIESKIEEISVWSKSKIEISSVSVSGKEQNHTEKQVAIET 438
QY 223 SSL-----RSDPKLWALGTVATGLIGLAATGIQALALTPEPDSPTTTPDAAASATETA 278
Db 439 PEQKQVKEDEHLNL-----QGSFIEESTKQPISSKPTSSPD-----MTDAA 481
QY 279 TRDQLTKRAFQNPQNKVNIDELGNATIPSGVLKDDVVANIEEQAKAAGEAKQQAIEENNA 338
Db 482 TGGRVSSSSFRDKILO-----NFSRSTI--DSFSNISK--KRNSEANDENDENTNL 530
QY 339 Q---AQKYDEQQAQKQFELKVSSGAGYGLSGALLGGIGVAVTAALHRKNQPVQETTT 395
Db 531 KIPPEKRRKQFVLQSKNILLVSS-----TEDSHEPVKVTED 567
QY 396 TTTTNTTTTTSARTVENKPNANTPA-----QGNVDTP-----GSEDPTMES---RRS 436
Db 568 SQTAHVSKFEDLENKSMESQSLQLLSESENDDKPLIDLIPLLAIKRNKDNLVSGVLEKG 627
QY 437 SMATSTSTFTFTSSI 451
Db 628 KSTSTSTKFTDTSIV 642
```

Search completed: September 27, 2001, 14:22:56
Job time: 152 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:27:03 ; Search time 38.06 Seconds
(without alignments)
1943.208 Million cell updates/sec

Title: US-09-189-415A-4

Perfect score: 2851

Sequence: 1 MFIGNLGHNPVNNNSIPPAP.....SNSAVNTSNPPAPGSHRFV 559

Scoring table: BLOSUM62 /

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:*

1: sp.archaea:*

2: sp.bacteria:*

3: sp.fungi:*

4: sp.human:*

5: sp.invertebrate:*

6: sp.mammal:*

7: sp.mhc:*

8: sp.organellae:*

9: sp.phage:*

10: sp.plant:*

11: sp.rodent:*

12: sp.unclassified:*

13: sp.vertebrate:*

14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Length | DB ID | Description |
|------------|--------|---------|--------|----------|--------------------|
| 1 | 2624.5 | 92.1 | 558 | 2 Q9R396 | Q9R396 escherichia |
| 2 | 2608.5 | 91.5 | 558 | 2 Q85506 | Q85506 escherichia |
| 3 | 1774.5 | 62.2 | 538 | 2 Q47014 | Q47014 escherichia |
| 4 | 1766.5 | 62.0 | 538 | 2 Q47016 | Q47016 escherichia |
| 5 | 1753.5 | 61.5 | 538 | 2 Q85508 | Q85508 escherichia |
| 6 | 1570 | 55.1 | 551 | 2 Q82258 | Q82258 escherichia |
| 7 | 1554.5 | 54.5 | 552 | 2 Q9KWH9 | Q9KWH9 escherichia |
| 8 | 1548.5 | 54.3 | 550 | 2 Q52147 | Q52147 escherichia |
| 9 | 1522 | 53.4 | 547 | 2 Q9WXX1 | Q9WXX1 escherichia |
| 10 | 1521 | 53.3 | 547 | 2 Q9ETI1 | Q9ETI1 citrobacter |
| 11 | 1495 | 52.4 | 549 | 2 Q50190 | Q50190 escherichia |
| 12 | 190 | 6.7 | 1323 | 5 Q9NHX4 | Q9NHX4 drosophila |
| 13 | 170 | 6.0 | 1229 | 5 Q94185 | Q94185 caenorhabdi |
| 14 | 165.5 | 5.8 | 2232 | 5 P91365 | P91365 caenorhabdi |
| 15 | 162.5 | 5.7 | 1291 | 5 Q77261 | Q77261 drosophila |
| 16 | 161 | 5.6 | 1300 | 5 Q9W5E0 | Q9W5E0 drosophila |
| 17 | 161 | 5.6 | 2586 | 5 Q9VTK8 | Q9VTK8 drosophila |
| 18 | 156 | 5.5 | 2921 | 5 Q9N973 | Q9N973 leishmania |
| 19 | 152.5 | 5.3 | 1412 | 5 Q9VKJ1 | Q9VKJ1 drosophila |

| | | | | | |
|----|-------|-----|------|-----------|--------------------|
| 20 | 149.5 | 5.2 | 676 | 5 Q9VB94 | Q9VB94 drosophila |
| 21 | 148.5 | 5.2 | 770 | 3 Q9P319 | Q9P319 neurospora |
| 22 | 148 | 5.2 | 1192 | 5 Q17346 | Q17346 caenorhabdi |
| 23 | 148 | 5.2 | 1192 | 5 Q9TW45 | Q9TW45 caenorhabdi |
| 24 | 147.5 | 5.2 | 3012 | 5 Q97205 | Q97205 leishmania |
| 25 | 145 | 5.1 | 332 | 5 Q9V8Q1 | Q9V8Q1 drosophila |
| 26 | 144.5 | 5.1 | 3507 | 5 Q23587 | Q23587 caenorhabdi |
| 27 | 144 | 5.1 | 1323 | 6 Q77509 | Q77509 bos taurus |
| 28 | 143.5 | 5.0 | 1275 | 5 Q76602 | Q76602 caenorhabdi |
| 29 | 143.5 | 5.0 | 1805 | 11 Q63661 | Q63661 rattus norv |
| 30 | 143 | 5.0 | 2678 | 5 Q9NDS4 | Q9NDS4 dictyostell |
| 31 | 142.5 | 5.0 | 1026 | 5 Q9W8B9 | Q9W8B9 drosophila |
| 32 | 141.5 | 5.0 | 1086 | 5 Q9W3U2 | Q9W3U2 drosophila |
| 33 | 141.5 | 5.0 | 1096 | 5 Q17368 | Q17368 caenorhabdi |
| 34 | 141.5 | 5.0 | 1144 | 5 Q9TVG6 | Q9TVG6 caenorhabdi |
| 35 | 141 | 4.9 | 909 | 13 Q9PTK7 | Q9PTK7 xenopus lae |
| 36 | 141 | 4.9 | 1475 | 5 Q9GYW7 | Q9GYW7 drosophila |
| 37 | 140.5 | 4.9 | 875 | 5 Q9W468 | Q9W468 drosophila |
| 38 | 140.5 | 4.9 | 1829 | 5 Q22248 | Q22248 caenorhabdi |
| 39 | 140 | 4.9 | 576 | 2 Q9KJ33 | Q9KJ33 enterococcu |
| 40 | 139 | 4.9 | 2468 | 2 Q9I2M3 | Q9I2M3 pseudomonas |
| 41 | 138 | 4.8 | 1038 | 10 Q9LDS2 | Q9LDS2 oryza sativ |
| 42 | 137.5 | 4.8 | 1385 | 5 Q9W4Z0 | Q9W4Z0 drosophila |
| 43 | 137.5 | 4.8 | 4498 | 5 Q9W2Z3 | Q9W2Z3 drosophila |
| 44 | 137 | 4.8 | 620 | 5 Q9VKL0 | Q9VKL0 drosophila |
| 45 | 137 | 4.8 | 837 | 5 P90666 | P90666 anthocidari |

ALIGNMENTS

| | |
|--|---|
| RESULT 1 | |
| Q9R396 | |
| ID Q9R396 | PRELIMINARY; PRT; 558 AA. |
| AC Q9R396; | |
| DT 01-MAY-2000 (TREMBlrel. 13, Created) | |
| DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update) | |
| DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update) | |
| DE TRANSLOCATED INTIMIN RECEPTOR TIR. | |
| GN TIR. | |
| OS Escherichia coli. | |
| OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; | |
| OC Escherichia. | |
| OX NCBI_TaxID=562; | |
| RN [1] | |
| RP SEQUENCE FROM N.A. | |
| RC STRAIN=86/24; | |
| RA MEDLINE=99242825; PubMed=10225900; | |
| RA Devlinney R., Stein M., Reinscheid D., Abe A., Ruschkowski S., | |
| RA Finlay B.B.; | |
| RT "Enterohemorrhagic Escherichia coli O157:H7 produces TIR, which is | |
| RT translocated to the host cell membrane but is not tyrosine | |
| RT phosphorylated."; | |
| RL Infect. Immun. 67:2389-2398(1999). | |
| RN [2] | |
| RP SEQUENCE FROM N.A. | |
| RC STRAIN=EDL933; | |
| RA MEDLINE=98339885; PubMed=9673266; | |
| RA Perna N.T., Mayhew G.F., Postai G., Elliott S., Donnenberg M.S., | |
| RA Kaper J.B., Blattner F.R.; | |
| RT "Molecular evolution of a pathogenicity island from enterohemorrhagic | |
| RT Escherichia coli O157:H7."; | |
| RL Infect. Immun. 66:3810-3817(1998). | |
| DR EMBL; AF125993; AAD29391.1; - | |
| DR EMBL; AF071034; AAC31506.1; - | |
| DR InterPro; IPR003536; - | |
| DR PRINTS; PR01370; TRNSINTIMNR. | |
| KW Receptor. | |
| Q9R396 | SEQUENCE 558 AA; 58022 MW; 99C417222D4B4A1 CRC64; |
| Query Match | 92.1%; Score 2624.5; DB 2; Length 558; |
| Best Local Similarity | 94.0%; Pred. No. 1e-152; |

| Matches 528; Conservative 3; Mismatches 24; Indels 7; Gaps 3; | | Matches 525; Conservative 3; Mismatches 27; Indels 7; Gaps 3; | |
|---|---|---|--|
| QY | 1 MPTGNLGNPNVNSIPPAPPLPSQTDGAGRGQLNSTGPGSLRALFTPVNRNSMADSGD 60 | QY | 1 MPTGNLGNPNVNSIPPAPPLPSQTDGAGRGQLNSTGPGSLRALFTPVNRNSMADSGD 60 |
| DB | 1 MPTGNLGNPNVNSIPPAPPLPSQTDGAGRGQLNSTGPGSLRALFTPVNRNSMADSGD 60 | DB | 1 MPTGNLGNPNVNSIPPAPPLPSQTDGAGRGQLNSTGPGSLRALFTPVNRNSMADSGD 60 |
| QY | 61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPDLTNRQIGSSVFRVETQEDGKHIA 120 | QY | 61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPDLTNRQIGSSVFRVETQEDGKHIA 120 |
| DB | 61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPDLTNRQIGSSVFRVETQEDGKHIA 120 | DB | 61 NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPDLTNRQIGSSVFRVETQEDGKHIA 120 |
| QY | 121 VGORNGVETSVLSDQYARLQSIDPEGDKKFVFTGGRGAGHAMVTVASDITEARQRIIL 180 | QY | 121 VGORNGVETSVLSDQYARLQSIDPEGDKKFVFTGGRGAGHAMVTVASDITEARQRIIL 180 |
| DB | 121 VGORNGVETSVLSDQYARLQSIDPEGDKKFVFTGGRGAGHAMVTVASDITEARQRIIL 180 | DB | 121 VGORNGVETSVLSDQYARLQSIDPEGDKKFVFTGGRGAGHAMVTVASDITEARQRIIL 180 |
| QY | 181 ELLEPKGTGSKGAGESKGVGELRESNSGAENTTETQSTSTSSLRSDPKLWLALGTVA 240 | QY | 181 ELLEPKGTGSKGAGESKGVGELRESNSGAENTTETQSTSTSSLRSDPKLWLALGTVA 240 |
| DB | 181 ELLEPKGTGSKGAGESKGVGELRESNSGAENTTETQSTSTSSLRSDPKLWLALGTVA 240 | DB | 181 ELLEPKGTGSKGAGESKGVGELRESNSGAENTTETQSTSTSSLRSDPKLWLALGTVA 240 |
| QY | 241 GLIGLAATGIVQALALTPEPDSPTTTPDPAASAATETATRDQLTKEAFQNPDKQKNIDE 300 | QY | 241 GLIGLAATGIVQALALTPEPDSPTTTPDPAASAATETATRDQLTKEAFQNPDKQKNIDE 300 |
| DB | 241 GLIGLAATGIVQALALTPEPDSPTTTPDPAASAATETATRDQLTKEAFQNPDKQKNIDE 300 | DB | 241 GLIGLAATGIVQALALTPEPDSPTTTPDPAASAATETATRDQLTKEAFQNPDKQKNIDE 300 |
| QY | 301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAENNAQAQKYDEQQAQRQBELKVSSG 360 | QY | 301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAENNAQAQKYDEQQAQRQBELKVSSG 360 |
| DB | 301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAENNAQAQKYDEQQAQRQBELKVSSG 360 | DB | 301 LGNAIPSGVLKDDVVANIEEQAKAAGEEAKQQAENNAQAQKYDEQQAQRQBELKVSSG 360 |
| QY | 361 AGYGLSGALLGGGIGVAVTAALHRKNQPVQETTTTTTTTTTSARTVENKPNANTPAQG 420 | QY | 361 AGYGLSGALLGGGIGVAVTAALHRKNQPVQETTTTTTTTTTSARTVENKPNANTPAQG 420 |
| DB | 361 AGYGLSGALLGGGIGVAVTAALHRKNQPVQETTTTTTTTTTSARTVENKPNANTPAQG 420 | DB | 361 AGYGLSGALLGGGIGVAVTAALHRKNQPVQETTTTTTTTTTSARTVENKPNANTPAQG 420 |
| QY | 421 NVDTPGSEDTMESRRSSMASTSTFFDTSIG---GPCRIRMLMKHRCMIRRCRLILI 477 | QY | 421 NVDTPGSEDTMESRRSSMASTSTFFDTSIG---GPCRIRMLMKHRCMIRRCRLILI 477 |
| DB | 421 NVDTPGSEDTMESRRSSMASTSTFFDTSIG---GPCRIRMLMKHRCMIRRCRLILI 477 | DB | 421 NVDTPGSEDTMESRRSSMASTSTFFDTSIG---GPCRIRMLMKHRCMIRRCRLILI 477 |
| QY | 478 RLFRWIGIQISVYVSTTQHPHPTDNGARLLGNPSAGIQSTYARLALSGGLRDMGGLT 537 | QY | 478 RLFRWIGIQISVYVSTTQHPHPTDNGARLLGNPSAGIQSTYARLALSGGLRDMGGLT 537 |
| DB | 480 QNM---GNTDSVYVSTTQHPHPTDNGARLLGNPSAGIQSTYARLALSGGLRDMGGLT 536 | DB | 480 QNM---GNTDSVYVSTTQHPHPTDNGARLLGNPSAGIQSTYARLALSGGLRDMGGLT 536 |
| QY | 538 GGSNSAVNTSNPPAPGSHRFV 559 | QY | 538 GGSNSAVNTSNPPAPGSHRFV 559 |
| DB | 537 GGSNSAVNTSNPPAPGSHRFV 558 | DB | 537 GGSNSAVNTSNPPAPGSHRFV 558 |
| RESULT 2 | | RESULT 3 | |
| ID | Q47014 PRELIMINARY; PRT; 558 AA. | ID | Q47014 PRELIMINARY; PRT; 538 AA. |
| AC | Q47014; | AC | Q47014; |
| DT | 01-NOV-1998 (TrEMBLrel. 08, Created) | DT | 01-NOV-1996 (TrEMBLrel. 01, Created) |
| DT | 01-NOV-1998 (TrEMBLrel. 08, Last sequence update) | DT | 01-MAY-1999 (TrEMBLrel. 10, Last sequence update) |
| DT | 01-MAR-2001 (TrEMBLrel. 16, Last annotation update) | DT | 01-MAR-2001 (TrEMBLrel. 16, Last annotation update) |
| DE | TRANSLOCATED INTIMIN RECEPTOR TIR. | DE | TRANSLOCATED INTIMIN RECEPTOR TIR. |
| GN | TIR. | GN | TIR. |
| OS | Escherichia coli. | OS | Escherichia coli. |
| OC | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; | OC | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; |
| OC | Escherichia. | OC | Escherichia. |
| OX | NCBI_TaxID=562; | OX | NCBI_TaxID=562; |
| RN | [1] | RN | [1] |
| RP | SEQUENCE FROM N.A. | RP | SEQUENCE FROM N.A. |
| RC | STRAIN=955F2; | RC | STRAIN=REPEC 84/110/1, AND E65/56; |
| RA | Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.; | RA | Krejan E.O.; |
| RT | "The translocated intimin receptors (Tir) of Shiga toxinogenic | RL | Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases. |
| RT | Escherichia coli isolates belonging to serogroups O26, O111, and O157 | RP | SEQUENCE FROM N.A. |
| RT | react with sera from patients with hemolytic-uremic syndrome and | RC | STRAIN=413/89-1 (O26:H-); |
| RT | exhibit marked sequence heterogeneity." | RC | MEDLINE=9829404; PubMed=9632251; |
| RL | Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases. | RA | Delbel C., Kraemer S., Chakraborty T., Ebel F.; |
| DR | EMBL; AF070087; AAC69314.1; . | RT | "Espe, a novel secreted protein of attaching and effacing bacteria, is |
| DR | InterPro; IPR003536; . | RT | directly translocated into infected host cells, where it appears as a |
| DR | PRINTS; PR01370; TRNSINTIMINR. | RL | tyrosine-phosphorylated 90 kDa protein."; |
| SQ | SEQUENCE 558 AA; 58175 MW; CA2CDDAC94527C2E CRC64; | RL | Mol. Microbiol. 28:463-474 (1998). |
| Query Match 91.5%; Score 2608.5; DB 2; Length 558; | | SEQUENCE FROM N.A. | |
| Best Local Similarity 93.4%; Pred. No. 9.5e-152; | | STRAIN=95ZG1; | |

RX MEDLINE=99003184; PubMed=9784578;
RA Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT "Translocated intimin receptors (Tir) of Shiga-toxicogenic Escherichia
coli isolates belonging to serogroups O26, O111, and O157 react with
sera from patients with hemolytic-uremic syndrome and exhibit marked
sequence heterogeneity.";
RL Infect. Immun. 66:5580-5586(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B10;
RA Marches O., Nougayrede J.-P., Boury M., Mainil J., Charlier G.,
RA Boullier S., De Rycke J., Milon A., Oswald E.;
RT "Role of Tir and Intimin in the pathogenesis of rabbit
enteropathogenic Escherichia coli.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; U59502; AAC32028.2; -;
DR EMBL; AJ223063; CAAL1065.1; -;
DR EMBL; AF070068; AAC69316.1; -;
DR EMBL; AF132728; AAD27868.1; -;
DR EMBL; AF113597; AAF03080.1; -;
DR InterPro; IPR003536; -;
DR PRINTS; PR01370; TRNSINTMINR.
KW Hypothetical protein; Receptor.
SQ SEQUENCE 538 AA; 55420 MW; 31D7A8E227B3D06C CRC64;

Query Match 62.2%; Score 1774.5; DB 2; Length 538;
Best Local Similarity 63.2%; Pred. No. 8.5e-101;
Matches 360; Conservative 61; Mismatches 106; Indels 43; Gaps 6;
QY 1 MPIGNLGNPNVNNSTIPAPPPLPSQTDGAGG-RGQLINSTGLSGRALFTPVNRNSMADSG 59
DB 1 MPIGNLGNPNVNNSTIPAPPPLPSQTDGAGGARNGLNSNGPMGSRLLFTPIRNSVADAA 60
QY 60 DNRA5DVPGLPNMRLAASEITLNDGFVLDHGPDLTLNRQIGSSVFRVETQEDGKH 119
DB 61 DSRASDIPGLTNPPLFAASEVSLHGALVLDHKGGLDTLNSAIGSSLFRVETRDGSHV 120
QY 120 AVQNGVETSVVLSQDEVARLOSIDPEKDKFVFTGGRGAGHAMVTVASDITEARQRI 179
DB 121 AIGQNGLETTTVVLSQDEFSLSQSLDPEGKNFVFTGGRGAGHAMVTVASDIAEARQRI 180
QY 180 LELLEPKGSGKAGESKGVGELRESNAGNTTETQSTSTSLRSDPKLWALGTVA 239
DB 181 IDKLEPKDKETKEPG-----DPNKGEGKIIEHTSTSTSLRADPKLWLSLGTIA 231
QY 240 TGLIGLAATGIQVALALTEPPDSPITTDPAASAATETATRDQLTKEAFQNPQNKVND 299
DB 232 AGLGMAATGIAQVALTEPPDDPITTDPAANAATAAQAADQLTKEAFQNPQNKVND 291
QY 300 ELGNAIPSGVLKDDVVVAIEEQAAGEAKQAQOAKYDEQAQKQEEELKVSS 359
DB 292 ENGNALPSGELKDDVVVAQIAEQAKAAGEAQAEATSNQAQKQYDEQAKREQEMLSS 351
QY 360 GAGYGLSGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTSARTVENKPNNTPAQ 419
DB 352 GVGYGISGALILGGGIGAGVTAALHRKNQPAEQITITRT-----VVDNPTNNAQA 403
QY 420 GNVDPGSEDTEMRSSMASTSTFTDTS-----IGGPCRIRMLKHKRCMTR 469
DB 404 GNTDTSPEESPASRRNSNASLNSGSDTSSTGCTVENPYADVMPNRDLSARIEPIYD 463
QY 470 RCLLLILRLFWIGIQISVVSTIQHPPTDTCARLLGNPSAGIQSTYARLALSGGL 529
DB 464 E-----VAADPNYSVQIHFSGNSPVGTG-RLVGTGPGGIGQSTYALLASSGGL 508
QY 530 RHDMMGLTGSGNSAVNTSNPPAPGSHRFV 559
DB 509 RLGMGLTGSGESAVSTANATPPGPARYV 538

RESULT 4
Q47016

ID Q47016 PRELIMINARY; PRT; 538 AA.
AC Q47016;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-AUG-1999 (TRENBLrel. 11, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=REPEC 83/39;
RA O'Brien R.A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=REPEC 83/39;
RA Krejany E.O.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=RDEC-1;
RA Agin T.S., Boedeker E.C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 172-538 FROM N.A.
RC STRAIN=REPEC RDEC-1;
RA Krejany E.O.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U59504; AAD19750.1; -;
DR EMBL; AF045568; AAC15683.1; -;
DR EMBL; U59503; AAB02941.1; -;
DR InterPro; IPR003536; -;
DR PRINTS; PR01370; TRNSINTMINR.
KW Receptor.
SQ SEQUENCE 538 AA; 55410 MW; 3132A969B7B3D06C CRC64;

Query Match 62.0%; Score 1766.5; DB 2; Length 538;
Best Local Similarity 63.0%; Pred. No. 2.6e-100;
Matches 359; Conservative 61; Mismatches 107; Indels 43; Gaps 6;
QY 1 MPIGNLGNPNVNNSTIPAPPPLPSQTDGAGG-RGQLINSTGLSGRALFTPVNRNSMADSG 59
DB 1 MPIGNLGNPNVNNSTIPAPPPLPSQTDGAGGARNGLNSNGPMGSRLLFTPIRNSVADAA 60
QY 60 DNRA5DVPGLPNMRLAASEITLNDGFVLDHGPDLTLNRQIGSSVFRVETQEDGKH 119
DB 61 DSRASDIPGLTNPPLFAASEVSLHGALVLDHKGGLDTLNSAIGSSLFRVETRDGSHV 120
QY 120 AVQNGVETSVVLSQDEVARLOSIDPEKDKFVFTGGRGAGHAMVTVASDITEARQRI 179
DB 121 AIGQNGLETTTVVLSQDEFSLSQSLDPEGKNFVFTGGRGAGHAMVTVASDIAEARQRI 180
QY 180 LELLEPKGSGKAGESKGVGELRESNAGNTTETQSTSTSLRSDPKLWALGTVA 239
DB 181 IDKLEPKDKETKEPG-----DPNKGEGKIIEHTSTSTSLRADPKLWLSLGTIA 231
QY 240 TGLIGLAATGIQVALALTEPPDSPITTDPAASAATETATRDQLTKEAFQNPQNKVND 299
DB 232 AGLGMAATGIAQVALTEPPDDPITTDPAANAATAAQAADQLTKEAFQNPQNKVND 291
QY 300 ELGNAIPSGVLKDDVVVAIEEQAAGEAKQAQOAKYDEQAQKQEEELKVSS 359
DB 292 ENGNALPSGELKDDVVVAQIAEQAKAAGEAQAEATSNQAQKQYDEQAKREQEMLSS 351
QY 360 GAGYGLSGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTSARTVENKPNNTPAQ 419
DB 352 GVGYGISGALILGGGIGAGVTAALHRKNQPAEQITITRT-----VVDNPTNNAQA 403
QY 420 GNVDPGSEDTEMRSSMASTSTFTDTS-----IGGPCRIRMLKHKRCMTR 469

```

Db 404 GNTDTSPEESPARRRNSNASLNGSDTSTGTVENPYADVGMPRNDSLARISEPIYD 463
Qy 470 RCRLLILRLFRINGIOISVYVSTIOHPRTDNDGARLLGNPSAGIQSTYARLALSGGL 529
Db 464 E-----VAADPNYSVIOHFGSNPVTG-RLVGTGQGIQSTYALLASSGGL 508
Qy 530 RHDMGGLTGGNSAVNTSNPPAPGSHRFV 559
Db 509 RLGMGLTGGGESAVSTANAAPTGPARFV 538

RESULT 5
ID O85508 PRELIMINARY; PRT; 538 AA.
AC O85508;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EPBC87A;
RA Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT "The translocated intimin receptors (Tir) of Shiga toxinogenic
RT Escherichia coli isolates belonging to serogroups O26, O111, and O157
RT react with sera from patients with hemolytic-uremic syndrome and
RT exhibit marked sequence heterogeneity.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF070069; AAC69318.1; -.
DR InterPro; IPR003536; -.
DR PRINTS; PRO1370; TRNSINTIMINR.
SQ SEQUENCE 538 AA; 55602 MW; 447052A0E3214D6D CRC64;

```

```

Query Match 61.5%; Score 1753.5; DB 2; Length 538;
Best Local Similarity 62.6%; Pred. No. 1.6e-99;
Matches 357; Conservative 61; Mismatches 109; Indels 43; Gaps 6;

Qy 1 MPTGNLGNPNVNNSTPPAPPLPSQTDGAGG--RGOLINSTGPLGSRALFTPVNRNSMADSG 59
Db 1 MPTGNLGNPNVNNSTPPAPPLPSQTDGAGG--RGOLINSTGPLGSRALFTPVNRNSMADSG 60
Qy 60 DNRSADVPGLPNMRLAASEITLNDGFEVLHDHGLDPLTLNRQIGSSVFRVETQEDGKH 119
Db 61 DSRASDIPGLPTNPLRFAASEVSLHGALVLDKGLDPLTLNRQIGSSVFRVETQEDGSHV 120
Qy 120 AVQORNGVETSVLSDQYARLOSIDPEGKDFVETGGRGGAGHAMVTVASDITEARORI 179
Db 121 AIGKRGLETVYLSDOEFSSQSLDPEGKDFVETGGRGGAGHAMVTVASDITEARORI 180
Qy 180 LELLEPKGTGESKAGSGVGLRESNSGAENT--TETQSTSLRSDPKLWALGTV 239
Db 181 IDKLEPKGTKEKPG-----DPNSGEGKIEIHTSTFSLRADPKLWLSLGTIA 231
Qy 240 TGLIGLAATGIVQALATPEPDSPTTTPDPAASATETATROLTKEAQNPDKNVND 299
Db 232 AGLGMAATGIAQVATPEPDDPTTTPDPAASATETATROLTKEAQNPDKNVND 291
Qy 300 ELGNAPSGVLKDDVYANIEEOAKAAGEAKQAQAIENNAQAQKYDEQAKROELKVS 359
Db 292 ENGNAPSGELIDVVAIAEQAKAAGEAKQAQAIENNAQAQKYDEQAKROELKVS 351
Qy 360 GAGYGLSGALILGGGIGVAVTAALHRKNQVPEQTITTTTTTTTTSARTVENKPNTPAQ 419
Db 352 GVGYGISGALILGGGIGVAVTAALHRKNQVPEQTITTTTTTTTTSARTVENKPNTPAQ 403
Qy 420 GNVDTPGSEDTMESRRSSMASTSTSTPDTSS-----IGGPCRIRMLMLKHCIRMIR 469

```

```

Db 404 GNTDTSPEESPARRRNSNASLNGSDTSTGTVENPYADVGMPRNDSLARISEPIYD 463
Qy 470 RCRLLILRLFRINGIOISVYVSTIOHPRTDNDGARLLGNPSAGIQSTYARLALSGGL 529
Db 464 E-----VAADPNYSVIOHFGSNPVTG-RLVGTGQGIQSTYALLASSGGL 508
Qy 530 RHDMGGLTGGNSAVNTSNPPAPGSHRFV 559
Db 509 RLGMGLTGGGESAVSTANAAPTGPARFV 538

RESULT 6
ID O68258 PRELIMINARY; PRT; 551 AA.
AC O68258;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98187918; PubMed=9529069;
RA Voss E., Paton A.W., Manning P.A., Paton J.C.;
RT "Molecular analysis of Shiga toxinogenic Escherichia coli O111:H-
RT proteins which react with sera from patients with hemolytic-uremic
RT syndrome.";
RL Infect. Immun. 66:1467-1472(1998).
DR EMBL; AF025311; AAC69249.1; -.
DR InterPro; IPR003536; -.
DR PRINTS; PRO1370; TRNSINTIMINR.
SQ SEQUENCE 551 AA; 56975 MW; 6EC95F76EF0F44CC CRC64;

```

```

Query Match 55.1%; Score 1570; DB 2; Length 551;
Best Local Similarity 59.5%; Pred. No. 2.7e-88;
Matches 339; Conservative 67; Mismatches 134; Indels 30; Gaps 11;

Qy 1 MPTGNLGNPNVNNSTPPAPPLPSQTDGA--GGRQLINSTGPLGSRALFTPVNRNSMADS 58
Db 1 MPTGNLGNPNVNNSTPPAPPLPSQTDGA--GGRQLINSTGPLGSRALFTPVNRNSMADS 60
Qy 59 GDNRSADVPGLPNMRL--AASEITLNDGFEVLHDHGLDPLTLNRQIGSSVFRVETQEDG 116
Db 61 VDSR--DIPGLPVHPSRLATATSEICLLGGFEVLHDHGLDPLTLNRQIGSARIEQSDG 118
Qy 117 KHIAVGQNGVETSVLSDQYARLOSIDPEGKDFVETGGRGGAGHAMVTVASDITEAR 176
Db 119 SYAAIGERNGVEVSVILNSQELQSLQIDIEDKGRFVETGGRGGGHSWVTPASDIAEAR 178
Qy 177 ORILELLEPKGTGESKAG---ESKGVGLRESNSGAENT--TETQSTSTSLRSDPKL 231
Db 179 AKILAKLPNNHGGSOARNVDRSVGVG---SASGMDSVVSETSTSTASSVRSRDPKF 234
Qy 232 WLALGTAVTGLIGLAATGIVQALATPEPDSPTTTPDPAASATETATROLTKEAFONP 291
Db 235 WYSIGAIAGLAGLAATGIVQALATPEPDSPTTTPDPAASATETATROLTKEAFONP 294
Qy 292 DNQKNIDELGNAPSGVLKDDVYANIEEOAKAAGEAKQAQAIENNAQAQKYDEQAKR 351
Db 295 ENQKVSIDEIGNISFSGELKDDVYAKIEEQAQEAQAVESNAQAQRYDYQYARR 354
Qy 352 QEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQVPEQTITTTTTTTTTSARTVENK 411
Db 355 QEELELSSGIGYGLSGALILGGGIGVAVTAALHRKNQVPEQTITTTTTTTTTSARTVENK 408
Qy 412 PANNTPAQGVNDTPGSEDTMESRRSSMASTSTSTF--DTSSIGGPCRIRMLMLKHCIRMIR 470
Db 409 TGGNTPAQGGTDAIRAEDTSLNRDSQRSTASTHWSDSSTAVVNYAEVGEARNSSPARQ 468

```


| | | | |
|----|-----|--|-----|
| Qy | 471 | CRLLILRLPRWIGIQISVYSTIQHPDRTDNGARLLGNFSAGIQSTYARLA-USGGI | 529 |
| | | : : | |
| Db | 469 | AEHHYDE-----VAADPNYSIQFNSGNQVGT-RLMGT PQGQSTYAILTNNASGL | 521 |
| Qy | 530 | RHDWGGITGGNSAVNTSNPPAPGSHREV | 559 |
| Db | 522 | RLGWGGITGGSGSAVNTANAAPTPGPGREV | 551 |

```

RESULT      7
Q9KWH9      PRELIMINARY;      PRT;      552 AA.
ID          Q9KWH9
AC          AC
Q9KWH9
DT          01-OCT-2000 (TEMBLrel. 15, Created)
DT          01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT          01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE          TIR.
GN          TIR.
OS          Escherichia coli.
OC          Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC          Escherichia
OX          NCBI_TaxID=562;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN=HK01;
RA          Abe A., Nagano H.;
RT          "Analyses of type III secreted proteins and Tir in enteropathogenic
RT          Escherichia coli O157:H45.";
RL          Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR          EMBL: AB036053; BAA96815.1; -.
DR          InterPro: IPR003536; -.
DR          PRINTS: PR01370; TRNSINTMINR.
SQ          SEQUENCE      552 AA; 57005 MW;      A8D79EE22EE50A4B CRC64;

```

| | | | |
|--------|---|--|--------------|
| Db | 461 | -----NSLSSTRQEEHHYDEVAADPVYSIQNFSRNAPVTG-RLMGSPGGQIG | |
| Qy | 518 | STYARLALSGGLRDMGGLTGGNSAVNT--SNNPPAPGSHREV | 559 |
| Db | 509 | STYALANSAGRLGMGGLTGGESAVNTANANAATPGVPREV | 552 |
| RESULT | 8 | | |
| ID | 052147 | PRELIMINARY; | PRT; 550 AA. |
| AC | 052147; | | |
| DT | 01-JUN-1998 | (T-EMBLrel. 06, Created) | |
| DT | 01-JUN-1998 | (T-EMBLrel. 06, Last sequence update) | |
| DT | 01-MAR-2001 | (T-EMBLrel. 16, Last annotation update) | |
| DE | TRANSLOCATED INTIMIN RECEPTOR. | | |
| DE | TIR. | | |
| OS | Escherichia coli. | | |
| OC | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; | | |
| OC | Escherichia. | | |
| OX | NCBI_TaxID=562; | | |
| RP | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K., | | |
| RA | Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.; | | |
| RL | Mol. Microbiol. 0:0-0(1998). | | |
| DR | EMBL; AF022236; AAC38390.1; -. | | |
| DR | InterPro; IPR003536; -. | | |
| DR | PRINTS; PR01370; TRNS1370. | | |
| SO | SEQUENCE 550 AA; 56509 MW. | | |
| SO | 19DD08A9BE9251CB CRC64; | | |

| | | | | | | | |
|-----------------------|--------|--------------|----------|------------|------|--------|------|
| Query Match | 53.48; | Score | 1522; | DB | 2; | Length | 547; |
| Best Local Similarity | 57.28; | Pred. No. | 2.3e-85; | | | | |
| Matches | 327; | Conservative | 68; | Mismatches | 139; | Indels | 38; |
| Gaps | | | | | | | |

| | | | | |
|----|-----|---|--|-----|
| QY | 1 | MPIGNLGNPNVNSIPAPLP | SDTDCG--GGRGQLINSTGLCSRALFTPVNRMSMADS | 58 |
| DB | 1 | MPIGNLGNNTSNLIIPAPLP | SDTATRGSSLLSSGSLGRLLFPPLRSSIYDT | 60 |
| | | : | : : : : : : | |
| QY | 59 | GNDRASDVPGLPVNPMRLAA | SEITLUNDGEVLVHDHGPLDTLNRQIGSSVFRVETQEDGKH | 118 |
| DB | 61 | VDSR--DVPGLPEHLRPA | TSETCLHGGEVLVHDKGPLDTLNKQIGSSVFRVETQEDGTH | 118 |
| | | : : : : : : : : : | : : : : : : : : : | |
| QY | 119 | TAVGQRNVETSVVLSDQYAR | LOSIDPEGKDKFFVTGRRGAGAHMTVASDITEAROR | 178 |
| DB | 119 | AAIGVKDGEVSVTLNSSL | ELQLOSLDTEGNRFVFTGRRGSGHAMTVASDTSQARE | 178 |
| | | : : : : : : : : : | : : : : : : : : : | |
| QY | 179 | IIELELEPKGTG----- | ESKGAGESKGVGELRESNGAENTTETQTSSTSLRSDP | 229 |
| DB | 179 | ITAKLDPDNHGRQPKDIT | RSVGVGSAGMGD-----GV--VSETHTSITVSSVRSDP | 230 |
| | | : : : : : : : : : | : : : : : : : : : | |
| QY | 230 | KLWLALGTVATGLIGLATG | IVOAALATPEPDSPTTDPDAASATETATRDQLTKEAFQ | 289 |
| DB | | : : : : : : : : : | : : : : : : : : : | |
| DB | 231 | KEWVSGLAAGLAGLATG | IVOAVALTPADPDITTPDEAANAEEAATKDQLTKEAFQ | 290 |
| | | : : : : : : : : : | : : : : : : : : : | |
| QY | 290 | NPDNQKVNITDELGNALP | SGVLKDDVVANIEEQAKAAGEAKQQAENNAQAOKKVEDQQA | 349 |
| DB | 291 | NPDNQKVNITDELGNALP | SGELKDDVYVAQIADQAQVAGEQARQAVESNAQAQRHDDQQA | 350 |
| | | : : : : : : : : : | : : : : : : : : : | |
| QY | 350 | KQOEELKVSSGAGYGLSG | ALILGGIGIVAVTAALHRKNOPVEQTITTTTTTTTSARTVE | 409 |
| DB | | : : : : : : : : : | : : : : : : : : : | |
| DB | 351 | KRQOEELDSSGIGYGLSS | ALIYGGGIGAGVTAMLHRNPPTQTTATHTS-----VLTQ | 403 |
| | | : : : : : : : : : | : : : : : : : : : | |
| QY | 410 | NKPANNTPAQGNVDTPG | SEBDMTESRSSMASTSST--FFDTS--SIGGPCRIRMLMLKHRCM | 467 |
| DB | 404 | QQTGGNTRAQGGADTGV | ENASILTRDSQASVASTOWSDTSGDVNPN-----YAEGWM | 456 |
| | | : : : : : : : : : | : : : : : : : : : | |
| QY | 468 | IRRCRLILILIRFIWGI | OISVVYSTIQHPHPTDNTDNGARLLGNPSAGISQTSYARLALSG | 527 |
| DB | 457 | SRNPSSLAAPEEP IY | DEVAPDPNYSYIOHFSGNNPYTG--RLVGSQGGQISQTYALLASSG | 515 |
| | | : : : : : : : : : | : : : : : : : : : | |

| Query Match | 53.3% | Score 1521 | DB 2 | Length 547 |
|-----------------------|--------------|-----------------------|---|------------|
| Best Local Similarity | 57.2% | Pred. No. 2.7e-85 | | |
| Matches 327 | Conservative | 68 | Mismatches 139 | Indels 38 |
| Gaps | | | | 10 |
| QY | 1 | MPIGNLGNPNVNNISIPAPLP | QSDGCA--GGRQLINSTGSLGRALFTPVNRSMADS | 58 |
| DB | 1 | MPIGNLGNNNISNLIIPAPLP | QSDGATRCNGSSLSSTGSLGRULLFSPLRSSIVDT | 60 |
| QY | 59 | GDNRASDVPGLFVNPMLRAAS | EITLNDGFEVLHDPGLDITLNRQIGSSVFRVETQEDGKH | 118 |
| DB | 61 | VDSR--DVPGLPEHPLRFAT | SETCLHGGEVLHDPGLDITLNRQIGSSVFRVETQEDGTH | 118 |
| QY | 119 | TAVGORNGETSVVLSDQEXA | RLOSIDPEGDKFVFTGGRGAGAHMTVASDITEAROR | 178 |
| DB | 119 | AAIGVKDQGEVSVTLNNS | LOQLSDTEGTGRFVFTGGRGSGAHMTVASDISOAREK | 178 |
| QY | 179 | ILLELEPKGTG----- | ESKGAGESKGVGELRESNGAENTTQTSTSTSLRSDP | 229 |
| DB | 179 | ITAKLDPDNHGRQPKD | IDITRSVGVGSASGMGD-----GV--VSETHSTTTSSVRSRP | 230 |
| QY | 230 | KLWLALGTVATGLIGLATG | IVQALATPEPDSPTTDDPAASATETATROOLKEAFQ | 289 |
| DB | 231 | KFWVSVAAGLAGLAATG | IVQAVALTAPDDPTTDPDEAANAEEATKQOLTKEAFQ | 290 |
| QY | 290 | NPDNQKVNIDELGNATIP | SGVLKDDVVYANTEEOAKAGEAKQQA IENNAOAKKYDEOQA | 349 |
| DB | 291 | NPDNQKVNIDELGNATIP | SGELKDDVVYQATDAQAGEAQARQAVESNAQQRHDDQQA | 350 |
| QY | 350 | KQOEELKVSSGAGYGLSG | ALILGGIGVATYAALHRKNQPVQETTTTTTTTTTSARTVE | 409 |
| DB | 351 | KQOEELDLSSGIGYGLSS | ALIYGGIGAGVTAMLHRNPPTQETIATTHS-----VIQ | 403 |
| QY | 410 | NKPANNTPAQGNVDTPG | SEDTMESRSSMASTSST--FFDTS--SIGGFCRIMLMKLRCH | 467 |
| DB | 404 | QQTGGNTRAGGADTTG | VENASLTRDSQASVASTQWSDTSGDVWNP-----YAEQWM | 456 |
| QY | 468 | IRRCRLILILIRPRING | IQISVYSTIQHPRTDTTNGARLLGNPSAGIOSTVARLALSG | 527 |
| DB | 457 | SRNPSELLAPEEP | IYDEVAPDPNYSYQHFSGNPNVTG--KLVSQGGIOSTVATLASSG | 515 |

THIS PAGE BLANK (USPTO)